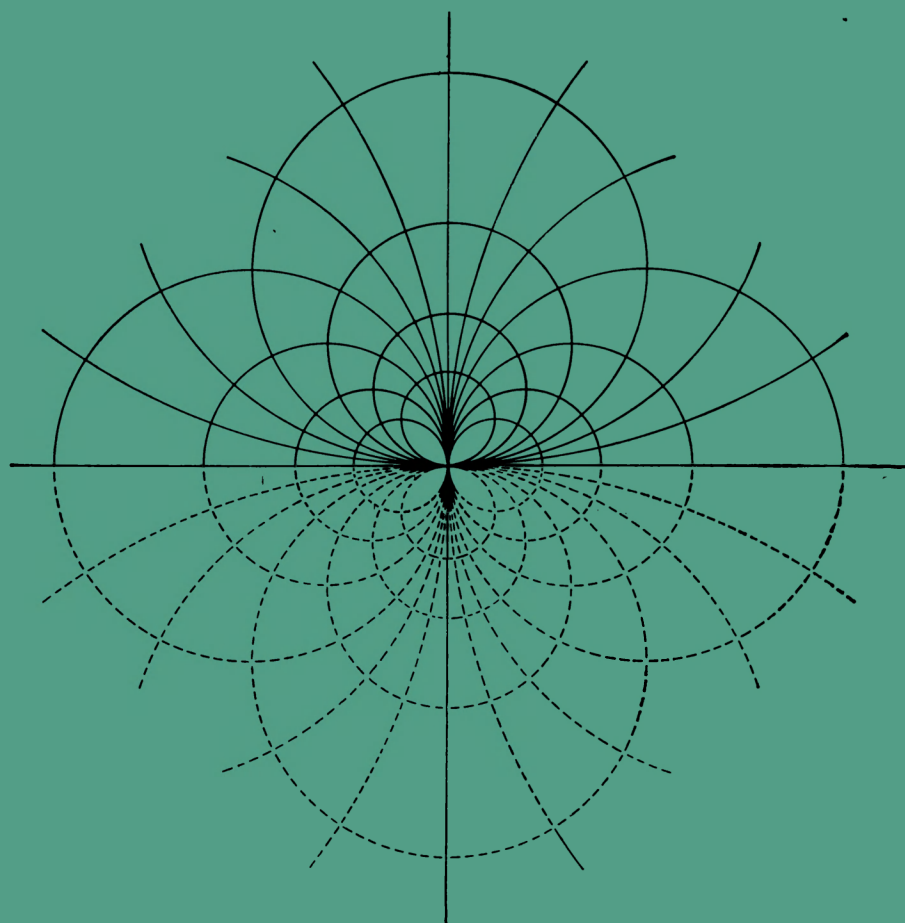


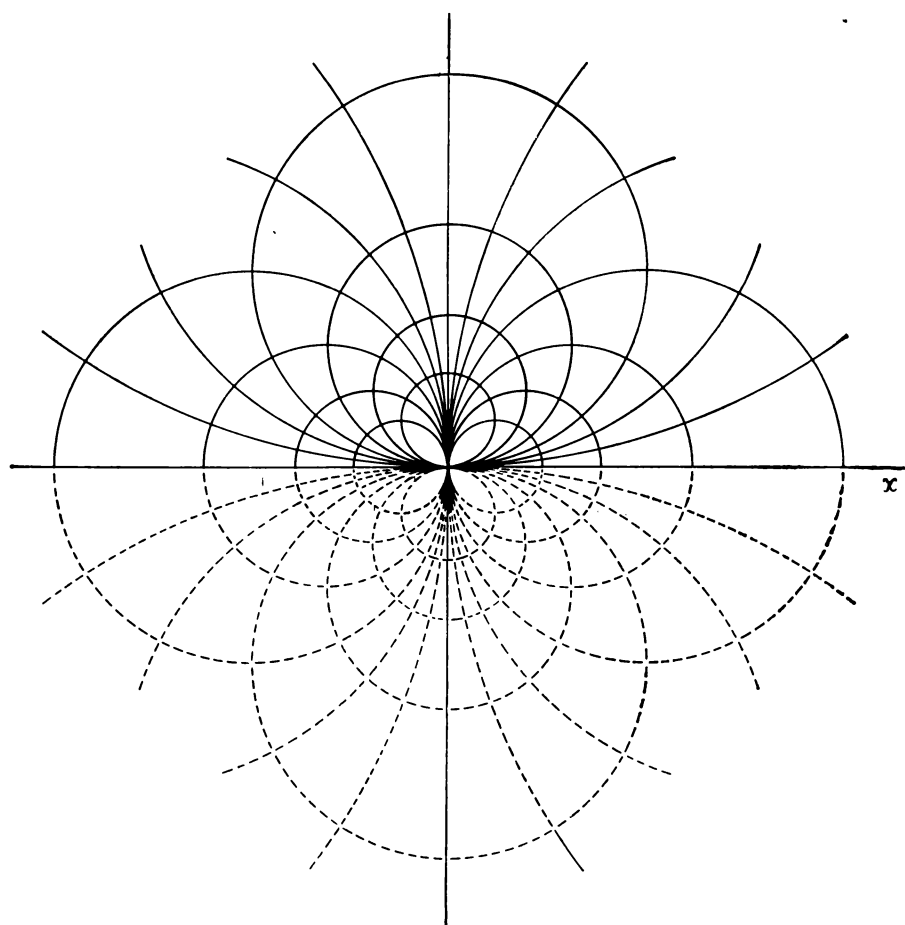
N. V. Efimov

Higher Geometry



Mir Publishers • Moscow





N. V. Efimov

Higher Geometry

Translated from the Russian
by
P. C. Sinha, Cand.Sc.

Mir Publishers
Moscow

Н. В. Ефимов

ВЫСШАЯ ГЕОМЕТРИЯ

Издательство «Наука»

First published 1980

Revised from the 1978 Russian edition

На английском языке

© English translation, Mir Publishers, 1980

Contents

Preface to the English Edition 9

Part 1. The Foundations of Geometry

<i>Chapter 1.</i>	A Short Review of Investigations into the Foundations of Geometry	12
1.1	Euclid's definitions, postulates and axioms	12
1.2	The fifth postulate	16
1.3	Lobachevsky and his geometry	33
1.4	The formation of the notion of geometrical space	36
<i>Chapter 2.</i>	The Axioms of Elementary Geometry	42
2.1	Geometrical elements	42
2.2	Group I: axioms of incidence	42
2.3	Group II: axioms of betweenness	45
2.4	Corollaries of the axioms of incidence and betweenness	45
2.5	Group III: axioms of congruence	53
2.6	Corollaries of Axioms I-III	57
2.7	Group IV: axioms of continuity	69
2.8	Group V: the axiom of parallelism. Absolute geometry	82
<i>Chapter 3.</i>	The Non-Euclidean Theory of Parallels	85
3.1	Lobachevsky's definition of parallels	85
3.2	Peculiarities in the location of parallel and divergent lines	96

3.3	Lobachevsky's function $\Pi(x)$	100
3.4	Lines and planes in Lobachevskian space	104
3.5	Equidistants and oricycles	111
3.6	Equidistant surfaces and horospheres	121
3.7	Elementary geometry on surfaces in Lobachevskian space	126
3.8	Area of a triangle	136
3.9	Proof of logical consistency of Lobachevskian geometry	146
3.10	The basic metric relationships in Lobachevskian geometry	165
3.11	Riemannian geometry: a brief survey	178

Chapter 4. Investigation of the Axioms of Elementary Geometry 188

4.1	Three basic problems of axiomatic theory	188
4.2	Consistency of the axioms of Euclidean geometry	191
4.3	Proof of independence of some axioms of Euclidean geometry	205
4.4	Axiom of completeness	216
4.5	The completeness of axioms of Euclidean geometry	220
4.6	The axiomatic method in mathematics	223

Part 2. Projective Geometry

Chapter 5. The Foundations of Projective Geometry 226

5.1	The subject matter of projective geometry	226
5.2	Desargues' theorem. Harmonic sets of elements	231
5.3	Order of points on the projective line	244
5.4	Division of harmonic pairs. Continuity of the harmonic correspondence	252
5.5	The axiom of continuity. Projective system of coordinates on a line	258
5.6	Projective coordinate systems on a plane and in space	270
5.7	Projective correspondence between elements of one-dimensional manifolds	283
5.8	Projective correspondences between two- and three-dimensional manifolds	292

- 5.9 Analytic representation for projectivities.
Involution 300
- 5.10 Transformation formulas for projective coordinates.
Cross ratio of four elements 316
- 5.11 The principle of duality 325
- 5.12 Algebraic curves and pencils. Algebraic surfaces
and bundles. Complex projective plane and complex
projective space 337
- 5.13 Images of second degree. The theory
of polars 346
- 5.14 Constructive theorems and problems of projective
geometry 362

Chapter 6. Group-Theoretic Principles in Geometry. Groups of Transformations 391

- 6.1 Geometry and the theory of groups 391
- 6.2 The projective group and its basic subgroups 395
- 6.3 Geometries of Lobachevsky, Riemann and Euclid
in the projective setting 408

Chapter 7. Minkowski Space 425

- 7.1 Multidimensional affine space 425
- 7.2 Euclidean spaces and Minkowski space 440
- 7.3 The space of events of the special relativity
theory 455

Part 3. Constant Curvature Geometry

Chapter 8. Differential Properties of Non-Euclidean Metric 474

- 8.1 Metric form of the Euclidean plane 474
- 8.2 The distance between two points on the Lobachevskian
plane 478
- 8.3 Metric form of the Lobachevskian plane 489
- 8.4 Intrinsic geometry of a surface and the Beltrami
problem 505
- 8.5 Geometry on a surface of constant curvature 511
- 8.6 Derivation of basic metric relations in Lobachevskian
geometry 523

<i>Chapter 9.</i>	Spatial Forms in Constant Curvature Geometry	529
9.1	Two-dimensional manifolds with differential-geometric metric	529
9.2	Parabolic spatial forms	537
9.3	Elliptic spatial forms	544
9.4	Hyperbolic spatial forms	547
	Name Index	553
	Subject Index	554

Preface to the English Edition

This book gives a systematic presentation of the fundamentals of Euclidean geometry, non-Euclidean geometry of Lobachevsky and Riemann, projective geometry and the geometrical aspects of special relativity theory. It also gives a general idea about constant curvature geometries. The subject matter has been shaped by the needs of the students and by my own experience.

The translation was done from the sixth Russian edition, which appeared in 1977. Extensive changes and additions were made in the 32 years that passed since the first Russian edition (1945). For instance, the chapter devoted to the Minkowski space and the special theory of relativity appeared in the fourth edition (1961). Also, I should like to point out the change in the formulation of Cantor's axiom. The first edition contained a somewhat stronger formulation of this axiom than is common. While preserving this formulation, the second edition contained an example of non-Archimedean geometry where Cantor's axiom is valid. However, this example is based only on the usual formulation of Cantor's axiom. This defect was pointed out to me by A. D. Aleksandrov. A number of other defects in the exposition has been corrected thanks to the review of my book by P. K. Rashevsky.

A few words on the use of the book as a textbook in a course of the foundations of geometry are in order. The subject matter is distributed over the first two parts. It is presented systematically throughout the book and practically does not omit any of the details of the arguments (except for the proofs of certain theorems of elementary geometry). It is evident that in lectures such a detailed presentation would be unprofitable (even if many hours were assigned to this course). The most difficult portion is Chapter 2 in the first part of the book; I feel that in lectures one should discuss from this chapter the formulations of the axioms and present examples of rigorous proofs of certain theorems; moreover, the most important aspects, for example, the measurement of length, the equivalence of Archimedes' and Cantor's axioms to Dedekind's axiom and the significance of these axioms for the substantiation of analytic geometry, should be discussed in detail. The proof of the majority of the initial theorems of elementary geometry should be left for self-study.

Chapter 3 (Lobachevskian geometry) can be lectured on with ease, and usually the students take an interest in the material. But here also many auxiliary theorems can be given without proof. In contrast, I feel that the logical consistency of Lobachevskian geometry must be studied in great detail as one of the most fundamental questions of the course (here one should use precise formulations of the axioms). As regards Chapter 4, one can be satisfied with a general review; it is enough to illustrate the concept of an incomplete axiomatic system by giving an example of axioms of the first group realized on a tetrahedron. If Chapter 4 is to be presented in some detail, I would recommend discussing non-Archimedean systems (with a short review of the main points of the Cartesian realization of all the axioms). In a course on foundations of geometry projective concepts must naturally play a secondary role. Accordingly, I feel that from Chapter 5 it is enough to take the introduction of projective coordinates on a line and on a plane (omitting the proof of the continuity of harmonic correspondence). I would recommend discussing Chapter 6 completely. The subject matter of the third part can be used at seminars.

Every new edition has benefited from questions, suggestions, comments and criticism from many students, colleagues and strangers. Among those to whom I owe some specific debt of gratitude are A. D. Aleksandrov, Nguyễn-Ca'hñ-Toàn (Vietnam), P. K. Rashevsky, I. A. Vainshtein and A. M. Zamorzaev.

Moscow, February 1980

N. V. Efimov

Part 1

The Foundations
of Geometry

Chapter 1

A Short Review of Investigations into the Foundations of Geometry

1.1. Euclid's definitions, postulates and axioms

§ 1. The origin of geometrical ideas can be traced back to distant past. The ancient cultures of Babylon and Egypt are usually credited with the first attempts at shaping them.

The period of development of geometry by the works of Greek scholars began in the seventh century B.C. Many basic facts pertaining to geometry were obtained in the sixth and fifth centuries B.C. and apparently the concept of the proof of a theorem had already been formulated by this time.

In the third century B.C. the Greeks already had a deep knowledge of geometry; not only had they accumulated a large number of geometric facts but they were also acquainted with the methods of geometric proofs. Therefore it was natural that in this period there were attempts to collect all these results together and put them in a logical order.

The task of describing the development of geometry was undertaken by many Greek authors, but their works have not survived to the present day. Apparently, all of them were forgotten after the appearance of Euclid's famous *Elements*.

§ 2. One of the greatest geometers of antiquity, Euclid lived about 300 B.C. His work *Elements* is divided into thirteen books, of which the fifth, seventh, eighth, ninth and tenth are devoted to the theory of proportion and to arithmetic (set forth in geometric form), while the remaining books are geometry proper.

The first book contains conditions for the equality of triangles, relationships between sides and angles of a triangle, the theory of parallel lines and conditions for triangles and polygons to be equal. The second book deals with the transformation of a polygon into square of equal area. The third book is devoted to circles, while the fourth studies inscribed and circumscribed polygons. The sixth book discusses the similarity of polygons. Foundations of solid geometry are set forth in the last three books.

Thus the *Elements* contain essentially elementary geometry. Much of that was already known in Euclid's time (e.g. the theory of conic sections) is not included there.

Euclid begins each of his books with definitions of those concepts that are required for the book in question.

The first book opens with twenty-three definitions. We cite here the first eight of them.

Definitions.

- (1) A point is that which has no part.
- (2) A line is breadthless length. [The word line means curve.]
- (3) The extremities of a line are points.
- (4) A straight line is a line which lies evenly with the points on itself.
- (5) A surface is that which has length and breadth only.
- (6) The extremities of a surface are lines.
- (7) A plane surface is a surface which lies evenly with the straight lines on itself.
- (8) A plane angle is the inclination to one another of two lines in a plane which meet one another and do not lie in a straight line.

After these definitions Euclid lists postulates and axioms, that is, assertions which are accepted without proof.*

Postulates. Let the following be postulated:

- (1) To draw a straight line from any point to any point.
- (2) To produce a finite straight line continuously in a straight line.
- (3) To describe a circle with any centre and radius.
- (4) That all right angles are equal.
- (5) That, if a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are less than the two right angles.

Axioms.

- (1) Things which are equal to the same thing are also equal to one another.
- (2) If equals be added to equals, the wholes are equal.
- (3) If equals be subtracted from equals, the remainders are equal.
- (4) If equals be added to unequals, the wholes are not equal.
- (5) If equals be doubled, the results are equal.
- (6) If equals be halved, the results are equal.
- (7) Things which coincide with one another are equal to one another.
- (8) The whole is greater than the part.
- (9) Two straight lines cannot enclose any space.

The origin of some of these axioms (4, 5, 6 and 9) is subject to some doubt. In some editions of the *Elements* the fourth and fifth postulates are treated as axioms; therefore the fifth postulate is sometimes

* The list of definitions, postulates and axioms is not identical in different editions of the *Elements*. The list given here is one which is widely used.

referred to as the eleventh axiom. It is not at all clear by what principles statements were classified as either postulates or axioms.

Following on from the axioms, Euclid sets forth theorems of geometry, arranging them in a logical order so that each proposition is proved using the previous propositions, postulates and axioms.

§ 3. The enumeration of definitions and axioms sufficient for the logical proof of all the following theorems is called (axiomatic) substantiation of geometry.

The problem of substantiation of geometry was posed by Euclid clearly in his *Elements* and was solved by him as accurately as was possible in his time. What is more, for many centuries the rigorous proofs of Euclid were something of a model to be imitated.

But if the exposition of the *Elements* is examined from the viewpoint of modern mathematics, it will be found that it is unsatisfactory in many respects.

First, let us examine Euclid's definitions, some of which have been mentioned above.

These definitions have been formulated in terms of concepts which themselves need to be defined, for example, "extremity", "length", etc. None of the Definitions 1-8 is used in the proof of any theorem; thus being unconnected with the remaining material of the *Elements*, they are essentially of no use and could be dropped altogether without any harm being done to the discussions which follow. These definitions are in fact only very naive descriptions of geometrical shapes.

As for the postulates and axioms, they are on the whole important; in the proof of many geometrical propositions one has to take into account the fact, for example, that a straight line is determined by two of its points, that a circle of arbitrary radius exists, etc. But here another feature must be pointed out: even a casual analysis shows that the list of basic statements, accepted by Euclid without proof, is so poor that it can hardly form a basis for a strictly logical development of geometry. To clarify this, let us cite some examples.

In geometrical arguments we frequently use concepts which we are accustomed to express in such phrases as "the given point of a line lies between two other points on the line", "two points lie on one side of the line", "two points lie on opposite sides of the line", "a point lies within a polygon", etc. Euclid's postulates do not give us any information with which to substantiate these concepts. When we use these statements to prove a theorem of some kind and we have only Euclid's postulates to help us, we are forced to rely on some visual evidence. But in a strictly logical construction any proposition not included in the axioms must be proved, no matter how evident it may seem to be.

It should be further noted that according to Axiom 7 the equality of geometrical quantities and figures is determined with the aid

of motion. But Euclid failed to define the concept of motion itself and the properties of motion are not mentioned in any of the axioms. Finally, whenever he considers two circles one of which passes through two points, one situated internally and one externally with respect to another circle, Euclid unreservedly assumes that these circles intersect; similarly, whenever he considers a line passing through a point situated within a circle, he takes it for granted that this line and the circle intersect at two points. Notwithstanding the fact that these results are self-evident, they must be proved. But there is no statement among the postulates and axioms of Euclid which might serve as a basis for such proofs.

Because of these facts, in many cases the convincing nature of Euclid's logic is corroborated by our habitual conception of space. This means that the *Elements* do not contain an unassailable logical substantiation of geometry.

§ 4. Some of the defects in Euclid's *Elements* were noted by the scholars of antiquity. In particular, the list of geometrical postulates was extended by Archimedes, who brought Euclid's exposition regarding the theory of lengths, areas and volumes to near completion. Whereas Euclid only established ratios of lengths, areas and volumes showing, for example, that the area of a circle is proportional to the square of its radius and the volume of a sphere is proportional to the cube of the radius, Archimedes obtained expressions that enable us to compute these quantities. To substantiate metric geometry, Archimedes introduces the following five postulates:

(1) Of all the curves having common end points, the straight line is the shortest.

(2) Two curves with common end points which are situated in the same plane are not equal if both of them are convex and one of them is enclosed by the other curve and by the straight line joining the end points; this holds true also if the curves have a common part and the remaining part is enclosed, the enclosed curve being smaller than the enclosing curve.'

(3) Similarly, of all the surfaces having the same plane perimeter, the plane is the smallest.

(4) Two surfaces having a common plane perimeter are not equal if both of them are convex and one of them (or a part thereof) is enclosed by the other surface and by the plane having the same perimeter, the enclosed surface being smaller than the enclosing surface.

(5) What is more, of two unequal curves, surfaces or bodies the larger of the two can become smaller than the quantity obtained by a suitable number of repetition of the smaller.

The first four propositions of Archimedes are unacceptable as postulates for the logical substantiation of metric geometry. They deal with the length of a curve, area of a surface and volume of a

body, but these concepts must be defined in terms of other simpler geometrical concepts. If these definitions are set forth in a proper form, the propositions of Archimedes can be proved. Therefore it is unnecessary to treat them as postulates.

On the other hand, the last assertion, which is sometimes simply referred to as Archimedes' postulate, is of great importance. It can be formulated in short as follows: for any two quantities a and b , $a < b$, there is an integer n such that $na > b$. This postulate is fundamental to the measurement of geometrical quantities, as will be shown in detail in Chapter 2, § 20.

After Archimedes, attempts to make the basic structure of geometry precise went on. But for many centuries nothing substantial in this connection was added to that had already been done by Euclid. Up until the nineteenth century the rigour of Euclid's proofs was generally thought to be sufficient. It was only at the end of the nineteenth century that the idea of a strictly logical construction of geometry was finally conceived and a complete system of axioms found from which all the theorems could be derived, without any recourse to our usual conception of space.

Very few geometers realized the necessity of extending Euclid's postulates; on the other hand, most of the scientists dealing with Euclid's *Elements* set themselves the task of reducing the number of statements accepted without proof. This expressed the natural desire to see with what minimum number of postulates and axioms geometry could be logically developed.

One result obtained in this direction without much difficulty was that Euclid's fourth postulate was seen to be superfluous, since the equality of right angles, like many other propositions, can be rigorously proved.

The majority of works on the foundations of geometry were an attempt to drop Euclid's fifth postulate from the list of basic assumptions because it appeared to be too complicated to be included as a postulate.

Investigations devoted to the fifth postulate are as old as Euclid's *Elements*. They were completed only towards the end of the nineteenth century and led to very important discoveries.

We shall now turn to the history of the fifth postulate; this will help to understand the contemporary problems of substantiating geometry.

1.2. The fifth postulate

§ 5. Every student of elementary geometry knows the fundamental role played by the fifth postulate: it forms the basis of the theory of parallel lines and all those problems connected with it, such as the similarity of figures and trigonometry.

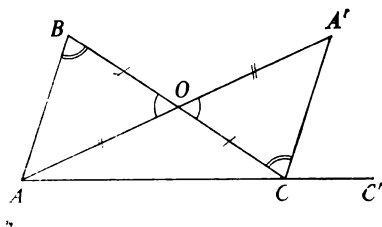


Figure 1

Let us refresh our memories by going over the sequence of initial propositions of plane geometry so as to demonstrate where the fifth postulate appears first.

School textbooks usually begin with the comparison of geometrical figures: segments, angles and triangles are considered equal if they can be made coincident through motion; a segment (angle) is greater than another segment (angle) if the latter can be made to cover a part of the former. But the concept of motion itself remains essentially undefined.

Next, some basic theorems are usually established. Among these are the theorems on the equality of triangles and on perpendicular and inclined lines and also the following theorems.

Theorem. *In an isosceles triangle the angles at the base are equal.*

Theorem. *An exterior angle of a triangle is greater than either remote interior angles.*

Theorem. *In a triangle the largest angle lies opposite the largest side (and conversely).*

Theorem. *Each side of a triangle is less than the sum of the other two sides.*

Of special interest to us is the theorem concerning interior and exterior angles of a triangle to which we shall frequently refer later. We present a proof of this theorem. Consider the triangle ABC (Fig. 1); we have to show that each of its exterior angles is greater than either remote interior angles. Let us try to establish this for the exterior angle at C and the interior angle at B .

Let O be the midpoint of the side BC . Join A and O and on the extension of AO take a point A' such that $AO = OA'$. Join A' and C and consider the two triangles AOB and $A'OC$. These two triangles are equal, since they contain equal angles subtended by two equal sides. From the equality of the two triangles it follows that $\angle ABC = \angle BCA'$. The proof of the theorem is completed, since $\angle BCA'$ is a part of the exterior angle at C .

The last step of the proof requires caution. The fact that $\angle BCA'$ is a part of $\angle BCC'$ or that point A' lies within $\angle BCC'$ (C' is any point on the extension of AC) is essentially proved simply by

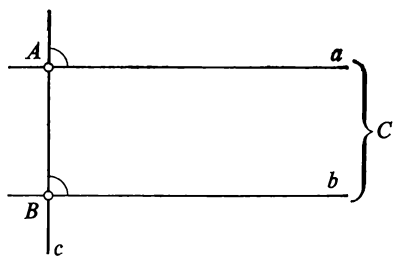


Figure 2

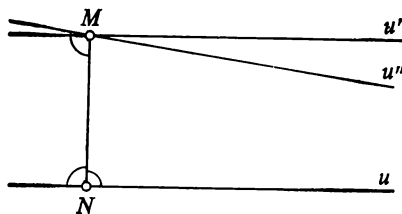


Figure 3

looking at the figure. As noted earlier, Euclid's axioms do not allow us to establish precisely the concepts "between", "within", etc.

What is more, the concept of equality of triangles which has been used above is not substantiated since Euclid does not define motion.

Thus the above arguments depend heavily on the visual evidence. Of course, similar remarks apply to the derivation of almost all geometric theorems. What is most important to note, however, is that the fifth postulate need be used to prove neither the theorem on interior and exterior angles of a triangle nor the abovementioned theorems.

Having established these propositions we give the definition of parallel lines: two lines are parallel if they do not have any point in common.*

For this definition to make sense the existence of parallel lines must be established. This can be done by means of a well-known theorem, which states that two lines perpendicular to a third line are parallel, the theorem being a direct consequence of the proposition regarding interior and exterior angles of a triangle.

Indeed, let two lines a and b make right angles with another line c at points A and B (Fig. 2). Assume that a and b are not parallel and let C be their point of intersection. Then the exterior angle of triangle ABC at A must be greater than the interior angle at B , which contradicts the assumptions made regarding these angles. This completes the proof of the theorem.

This readily implies that through any point M a line may be drawn parallel to a given line u not passing through this point (Fig. 3). To do this, we simply draw a perpendicular MN from M to u and draw a line u' perpendicular to MN at M . From our previous example we may conclude that u' will be parallel to u .

Having established the existence of parallel lines and having shown that through any point a line may be drawn parallel to a given

* We would like to remind our reader that we are now considering plane geometry.

line, we are left with one more question to be settled; namely, as to whether through a given point in the plane there passes only one line parallel to the given line or if there are many such lines.

In the theory of parallel lines it is proved that *through any point not lying on a given line there passes only one line parallel to the given line*. We present a proof of this fact (Fig. 3).

Let u be any line and M a point not on this line, and let MN be perpendicular to u . Denote by u' the line perpendicular to MN at M . We already know that u' is parallel to u . We draw through M another line u'' that does not coincide with u' and then show that u'' cannot be parallel to u . Since u'' does not coincide with u' , it forms an acute angle with MN on one of the sides. Thus u and u'' by intersecting MN form interior angles on the same side, the sum of the angles being less than two right angles. By the fifth postulate this implies that u and u'' intersect.

We find that in the above proof of the uniqueness of parallel lines considerable use has been made of the fifth postulate. It is easy to see that conversely the fifth postulate can be proved as a theorem if it is assumed that through every point not on a given line there passes only one line parallel to the given line.

Indeed, let lines a and b (Fig. 4) form interior angles, whose sum is less than $2d$, on the same (given) side of a transversal c . We have to show that on this side of c lines a and b have a common point.

Let a and b form angles α and β with c and let, in accordance with the above condition,

$$\alpha + \beta < 2d \quad (*)$$

Assume further that γ is the angle adjacent to α . Through the point of intersection of a and c draw a line a' forming angle $\gamma' = \beta$ with c . Then a' and b are parallel. Indeed, assuming that a' and b meet, we contradict the theorem on interior and exterior angles of a triangle. Taking the uniqueness of a parallel line as a postulate, we

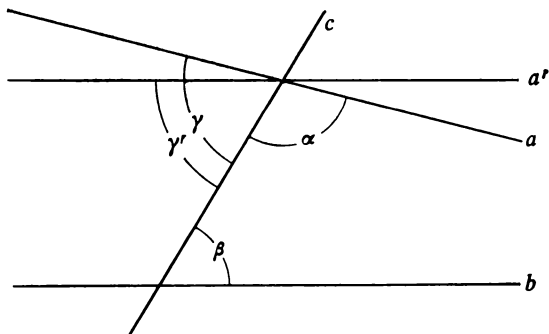


Figure 4

conclude that line a (not coinciding with a') is not parallel to b . It remains only to be proved that a and b meet on the side where angles α and β are situated. To do this note that $\alpha + \gamma = 2d$; this together with (*) implies that $\gamma > \beta$. Therefore a and b cannot meet on the side where γ lies, since otherwise γ would be an interior and β an exterior angle of the resulting triangle and the inequality $\gamma > \beta$ would be violated.

Thus the fifth postulate is equivalent to the assertion that there exists only one line passing through a given point and parallel to a given line; the entire Euclidean geometry is based on this fact. This implies, in particular, that two parallel lines form equal corresponding angles when they intersect a third line, that the sum of the interior angles of a triangle equals two right angles, and many other theorems. Thus the fifth postulate or, as it is also known, the parallel postulate forms the basis for the majority of the important propositions of elementary geometry.

§ 6. It is probable that Euclid tried to prove the parallel postulate. As evidence of this, we find that the first twenty-eight propositions of the *Elements* do not use the fifth postulate, as if Euclid tried to avoid using it as long as possible.

From the time of Euclid until the end of the nineteenth century the problem of the fifth postulate was one of the most popular problems of geometry. During this period several proofs of this postulate were put forward but all turned out to be erroneous. Usually the authors of these proofs used some geometrical assertion which was so obvious that it slipped into the proof unnoticed. At the same time attempts to establish any such assertion without using the fifth postulate proved to be unsuccessful.

Of course, such investigations did not attain their goals, since the problem consisted in freeing the Euclidean parallel theory from this special postulate. Hence the task here was not so much to replace the fifth postulate by some other assertion, even a very obvious one, but to prove this postulate using the remaining postulates of geometry.*

It should be noted, however, that despite their futility many attempts to prove the fifth postulate led to some very valuable results. It was only due to these attempts that the logical interdependence of various geometrical statements was discovered; in particular, several statements equivalent to the Euclidean parallel postulate were found (i.e. statements accepted without proof that make it possible to prove the fifth postulate with the aid of the other basic facts of Euclidean geometry).

The following are examples of propositions equivalent to the fifth postulate.

* This problem will be dealt with more precisely below.

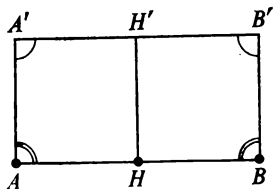


Figure 5

- (1) Through any point not lying on a given line there passes only one line parallel to the given line.
- (2) If a line cuts two parallel lines, the corresponding angles are equal.
- (3) The sum of the interior angles of a triangle is equal to two right angles.
- (4) Points lying at equal distances from a line on the same side of the line form a straight line.
- (5) The distances between the points on one of two parallel lines and those on the second are totally bounded.
- (6) There are triangles with arbitrarily large areas.
- (7) There are similar triangles.

Any of these propositions can be taken as the base of the theory of parallel lines, that is, assuming any one of them to be obviously true, one can prove the fifth postulate by precise argument and then, following Euclid, from it derive all other theorems. That the fifth postulate is equivalent to the abovementioned as well as to some other propositions will be proved later.

§ 7. From among a large number of works devoted to the fifth postulate particular mention must be made of those of the Jesuit priest and Italian mathematician Girolamo Saccheri (1667-1733) and the Swiss-German writer on a wide variety of mathematical and nonmathematical themes Johann Heinrich Lambert (1728-1777). Both made a considerable contribution to the task of substantiating the theory of parallel lines.

In the very year in which Saccheri died, he published a book entitled *Euclides ab omni naevo vindicatus* (Euclid Cleared of Every Flaw). In this book Saccheri tries to prove the fifth postulate by the method of *reductio ad absurdum*.

Saccheri begins by considering a quadrilateral $AA'B'B$ (Fig. 5) with two right angles at the base AB and two equal lateral sides AA' and BB' . From the symmetry of the figure about the perpendicular HH' to the midpoint of AB it follows that the angles at the vertices A' and B' are equal. If the fifth postulate is assumed to be true, and hence the Euclidean parallel theory, then it can readily be shown that angles A' and B' are right angles and $AA'B'B$ is a rectangle. Con-

versely, as shown by Saccheri, if even in one such quadrilateral the summit angles are right angles, Euclid's parallel postulate will hold. In order to prove this postulate, Saccheri proposes three possibilities: angles A' and B' are either right or obtuse or acute. He calls these three assumptions the right, the obtuse and the acute angle hypothesis, respectively. Since the right angle hypothesis is equivalent to the fifth postulate, to prove that postulate the remaining two possibilities must be shown to be unacceptable. Through precise argument Saccheri first of all shows that the obtuse angle hypothesis leads to a contradiction. Next, assuming the acute angle hypothesis to hold, he obtains important results, but arrives at two contradictory statements. Developing these results further, Saccheri constructs a complicated geometrical system some of whose propositions stand in such blatant contradiction to our own clear conception of how lines are located in a plane that they could be considered absurd. For example, in the geometrical system resulting from the acute angle hypothesis two parallel lines have either only one common perpendicular on either side of which they diverge limitlessly, or else they do not have any common perpendicular and, while converging asymptotically in one direction, they diverge limitlessly in the other.

Justifiably, Saccheri does not regard these propositions as logically inadmissible solely because they contradict our usual conception of space. Nevertheless, after a number of precise arguments Saccheri asserts the incorrectness of the acute angle hypothesis on the basis that two lines which become closer asymptotically must have a common perpendicular at the point of infinity, which is "contrary to the nature of a line". Thus supposing that the obtuse angle and the acute angle hypothesis lead to a contradiction, Saccheri concludes that the right angle hypothesis is correct and the proof of the fifth postulate is complete. Evidently Saccheri feels that he has not fully succeeded in showing that the acute angle hypothesis leads to a logical contradiction and he returns to it in order to prove that "this [hypothesis] contradicts itself". To this end, he calculates the length of a certain line in two ways and obtains different answers. This would have proved conclusively the desired contradiction had Saccheri not obtained his result by making a computational error.

The ideas developed by Lambert in his *Die Theorie der Parallelinien* (Theory of Parallel Lines) (1766) are similar to those of Saccheri.

Lambert begins by considering a quadrilateral $ABCD$ with three right angles at A , B and C (Fig. 6). Regarding the fourth angle there can be three alternatives: this angle is acute, right or obtuse. Here again we have three hypotheses. Having established that the right angle hypothesis is equivalent to the fifth postulate and shown that the obtuse angle hypothesis leads to a contradiction, Lambert, like

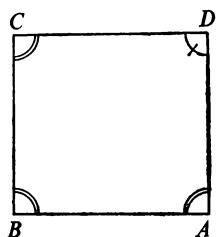


Figure 6

Saccheri, is forced to direct most of his attention to the acute angle hypothesis. Here again this hypothesis leads to a complicated geometrical system. However, in spite of the fact that this system was developed to a great extent by him, Lambert was not able to find any logical contradiction. The paradox concerning the location of lines in a system based on the acute angle hypothesis, which was noted above in connection with Saccheri's works, is also found in the works of Lambert. Like Saccheri, Lambert did not conclude the inadmissibility of the acute angle hypothesis solely on the basis that it contradicts the known properties of lines. But unlike Saccheri, Lambert did not commit any error by which he might consider the acute angle hypothesis rejected and hence the fifth postulate proved. Nowhere in his works does Lambert assert that he has proved the fifth postulate and he makes the firm conclusion that all other efforts to find a proof have come to nought:

Proofs of the Euclidean postulate can be developed to such an extent that apparently a mere trifle remains. But careful analysis shows that in this seeming trifle lies the crux of the matter; usually it contains either the proposition that is being proved or a postulate equivalent to it.

What is more, in the process of developing the system resulting from the acute angle hypothesis, Lambert discovers an analogy with spherical geometry and in this sees the possibility of the hypothesis being actually true:

I am even inclined to think that the third hypothesis is valid on some imaginary sphere. There must be some reason why it is difficult to reject it for the plane, as can be easily done with the second hypothesis.

As we shall see later, Lambert had a remarkable presentiment concerning the solution of the fifth postulate problem. In all events, not one of his predecessors made as much progress toward a correct solution as did Lambert.

§ 8.⁵ We shall now examine the researches of the French mathematician Adrien Marie Legendre (1752-1833), who is well known for his works in analysis and mechanics and who also made an important contribution to geometry.

Over a long period of time, Legendre tried to prove the fifth postulate and published several versions of the "proof". Although none of the versions proved to be correct, Legendre's researches are of interest, for they explain the connection between the fifth postulate and the proposition on the sum of the interior angles of a triangle.

In Euclidean geometry the proof of the theorem that the sum of the interior angles of a triangle is equal to two right angles is widely known to be based on the fifth postulate.

Legendre begins by showing that, conversely, if the statement that the sum of the interior angles of a triangle equals two right angles is accepted without proof, the fifth postulate can be proved as a theorem.

Next, in order to prove the fifth postulate without introducing new postulates, Legendre examines three mutually exclusive possibilities:

- (1) The sum of the angles of a triangle is greater than two right angles.
- (2) The sum of the angles of a triangle is two right angles.
- (3) The sum of the angles of a triangle is less than two right angles.

By way of precise argument, Legendre shows that the first possibility leads to a contradiction. If he were successful in showing that the third possibility also leads to a contradiction without using the fifth postulate, he would have established that the sum of the angles is two right angles and hence also the fifth postulate. But in trying to show that the third possibility does indeed lead to a contradiction, he unknowingly uses an equivalent version of the fifth postulate.

Positive results obtained by Legendre are contained in the following propositions.

Proposition 1. *If the sum of the angles of any triangle is two right angles, the fifth postulate holds.*

To prove this let us consider a line a and a point A not situated on it (Fig. 7).

Let AB be a perpendicular drawn from A to a . We know that a line a' passing through A and perpendicular to AB does not intersect a . We must show that any other line passing through A intersects a . According to the assumption made, we can use the fact that the sum of the angles of any triangle is two right angles.

Let b be any other line passing through A and β the acute angle that this line makes with AB . Let us show that b intersects a on the

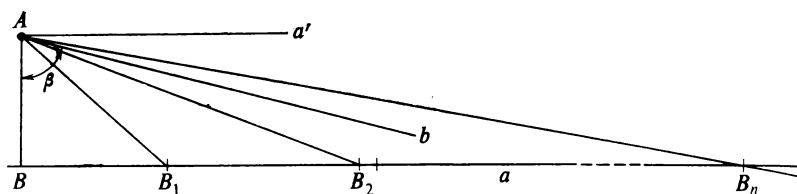


Figure 7

side of the acute angle. To do this, we take a point B_1 on a on the side of the acute angle so that the length BB_1 is equal to the length AB . On the same side as B_1 take another point B_2 so that B_1B_2 equals AB_1 and so on. Finally, take a point B_n so that $B_{n-1}B_n$ equals AB_{n-1} .

Consider triangles ABB_1 , AB_1B_2 , $\dots$, $AB_{n-1}B_n$. Since the sum of the angles in each of the triangles is equal to two right angles, in the isosceles triangle ABB_1 the interior angles at A and B_1 are each $\pi/4$.

Further, the interior angle at B_1 in triangle ABB_1 is exterior to triangle AB_1B_2 , and since the latter triangle is also isosceles, its two interior angles remote to B_1 are equal. From our hypothesis about the sum of angles of a triangle it follows that an exterior angle of a triangle is equal to the sum of the two remote interior angles. Therefore each of the interior angles in triangle AB_1B_2 at A and B_2 is $\pi/8$. Continuing in this way, we find that the interior angle at B_n in triangle $AB_{n-1}B_n$ is

$$\frac{1}{2^n} \frac{\pi}{2}$$

This implies that

$$\angle BAB_n = \frac{\pi}{2} - \frac{1}{2^n} \frac{\pi}{2}$$

Since β is acute, we may put

$$\beta = \frac{\pi}{2} - \varepsilon$$

with $\varepsilon > 0$. Choosing n so large that

$$\frac{1}{2^n} \frac{\pi}{2} < \varepsilon$$

we have $\beta < \angle BAB_n$.

In this case line b passes between the sides AB and AB_n of triangle BAB_n and consequently must intersect a' at a point that lies between B and B_n . ■ **

* A rigorous proof of this result can be given with the aid of Pasch's axiom (see § 13).

** Halmos' ■ is used to indicate the conclusion of a proof.

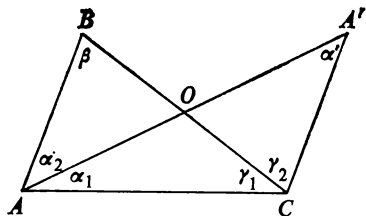


Figure 8

We shall now examine the question of possible values of the sum of the interior angles of a triangle. For convenience, we let $S(\triangle)$ denote the sum of the interior angles of triangle ' $\triangle$ ' and by $D(\triangle)$ the difference between two right angles and the sum of the interior angles so that

$$D(\triangle) = \pi - S(\triangle)$$

(this difference is called the *defect of a triangle*).

Proposition 2. *In every triangle*

$$S(\triangle) \leq \pi$$

The proof depends upon the following two lemmas.

Lemma 1. *In every triangle the sum of two interior angles is less than two right angles.*

Lemma 2. *For any given triangle another triangle can be constructed in which the sum of the angles is the same as that in the given triangle and in which one of the angles is at least half of any preassigned angle of the given triangle.*

We shall prove these lemmas.

The first lemma is a direct consequence of the proposition regarding interior and exterior angles of a triangle. Indeed, let α and β be interior angles of the given triangle and let α' be the exterior angle adjacent to α . Then

$$\alpha + \alpha' = \pi$$

But the exterior angle of a triangle is greater than the interior angle not adjacent to it. (The reader perhaps remembers that this proposition was proved without using the fifth postulate.) Therefore

$$\alpha' > \beta$$

and consequently

$$\alpha + \beta < \pi$$

To prove the second result, we consider any triangle ABC and show that another triangle can be constructed in which the sum of the angles is the same as in the given triangle and in which one angle is at least half of, for example, the angle at A in the given triangle (Fig. 8).

Denoting the midpoint of BC by O , we join A with O and extend AO to A' so that $AO = OA'$. Then $\triangle AA'C$ will have the required properties. In fact, using the notation of Fig. 8, we have

$$\begin{aligned} S(ABC) &= \alpha_1 + \alpha_2 + \beta + \gamma_1 \\ S(AA'C) &= \alpha_1 + \alpha' + \gamma_1 + \gamma_2 \end{aligned}$$

Since triangles ABO and COA' are evidently equal,

$$\alpha' = \alpha_2, \quad \gamma_2 = \beta$$

This implies first of all that the sum of the angles in the two triangles ABC and $AA'C$ is equal.

Next, the sum of the interior angles at A and A' in the second triangle is equal to the interior angle at A in the first triangle. Therefore one of them is at least half of the angle A in the triangle ABC , as required.

We shall now prove the main proposition of the lemma. This will be done by contradiction.

Assume that in some triangle $\triangle$ the sum of the angles is greater than two right angles so that $S(\triangle) = \pi + \varepsilon$, where $\varepsilon > 0$.

Let α denote one of the interior angles of $\triangle$. By Lemma 2, we can construct another triangle $\triangle_1$ one of whose interior angles α_1 will be at least half of α and $S(\triangle_1) = S(\triangle)$. We construct another triangle $\triangle_2$, in which one of the interior angles α_2 will be at least half of α_1 and $S(\triangle_2) = S(\triangle_1)$. Continuing in this manner, we construct a triangle $\triangle_n$ in which one of the interior angles α_n will be at least half of α_{n-1} and $S(\triangle_n) = S(\triangle_{n-1})$. Thus

$$S(\triangle_n) = \pi + \varepsilon \quad \text{and} \quad \alpha_n \leq \alpha/2^n$$

Choosing n so large that $\alpha/2^n < \varepsilon$ we find that $\alpha_n < \varepsilon$. But then the sum of the remaining two interior angles will be greater than π , which contradicts Lemma 1. ■

Thus, without using the fifth postulate, it can be asserted that the sum of the interior angles in a triangle does not exceed two right angles.

This feature will prove to be of immense importance later.

Following Legendre, we shall now show, without using the fifth postulate, that if there is at least one triangle in which the sum of the interior angles is assumed to equal two right angles, then the same is the case with any other triangle. But first let us establish the following lemmas.

Lemma 3. *If a transversal BP divides a triangle ABC into two triangles then the defect of ABC is equal to the sum of defects of triangles ABP and BPC .*

The proof can be perceived immediately. Indeed, following the notation of Fig. 9,

$$D(ABP) = \pi - (\alpha + \beta_1 + \delta_1)$$

$$D(BPC) = \pi - (\beta_2 + \delta_2 + \gamma)$$

giving

$$\begin{aligned} D(ABP) + D(BPC) &= 2\pi - (\alpha + \beta_1 + \beta_2 + \delta_1 + \delta_2 + \gamma) \\ &= \pi - (\alpha + \beta_1 + \beta_2 + \gamma) = D(ABC) \end{aligned}$$

Lemma 4. Suppose that two triangles ABC and AB_1C_1 are given in which A is the common vertex and vertices B_1, C_1 of the second triangle lie respectively on the sides AB and AC of the first. Then the defect of the second triangle does not exceed the defect of the first (Fig. 10).

The proof is an immediate consequence of Proposition 2 and Lemma 3. Indeed, joining points B and C_1 , by the Lemma 3 we obtain

$$D(ABC) = D(AB_1C_1) + D(B_1BC_1) + D(BC_1C)$$

But by Proposition 2, the defect of any triangle is either a positive number or zero. This together with the above equation yields

$$D(AB_1C_1) \leq D(ABC)$$

Lemma 5. Let two right angled triangles ABC and $A'B'C'$ be given, where the legs AC and BC of $\triangle ABC$ are, respectively, greater than the legs $A'C'$ and $B'C'$ of $\triangle A'B'C'$. Then, if the sum of the interior angles of $\triangle ABC$ is two right angles, the same is the case with $\triangle A'B'C'$.

To prove this lemma, we move $\triangle A'B'C'$ so that vertex C' coincides with C while legs $A'C', B'C'$ coincide with AC and BC , respectively. Then, by Lemma 4,

$$D(A'B'C') \leq D(ABC)$$

But $D(ABC) = 0$ by hypothesis and $D(A'B'C') \geq 0$ according to Proposition 2, therefore the previous inequality implies immediately that $D(A'B'C') = 0$.

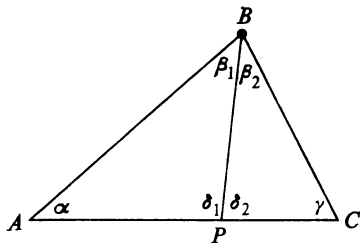


Figure 9

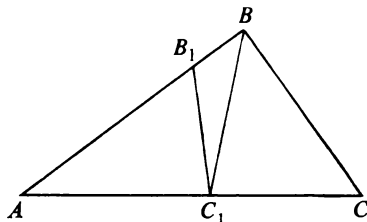


Figure 10

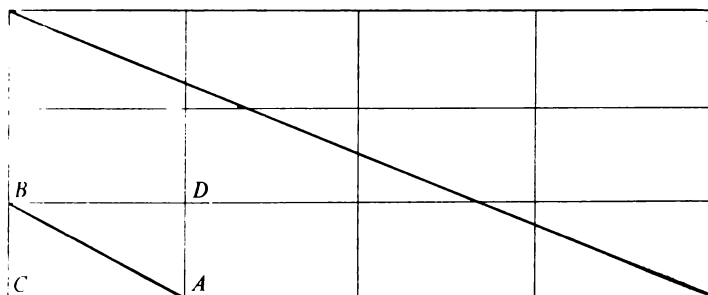


Figure 11

Lemma 6. *If the sum of the interior angles of a right triangle is equal to two right angles, then so is the sum of the angles of any other right triangle.*

Consider two right triangles ABC and $A'B'C'$. Suppose that the sum of the angles of $\triangle ABC$ is equal to two right angles. We have to show that the sum of the angles of $\triangle A'B'C'$ is also equal to two right angles. If the legs AC and BC of the first triangle are respectively longer than the legs $A'C'$ and $B'C'$ of the second triangle, then the assertion of the lemma is implied by Lemma 5. But if at least one of the legs of $\triangle ABC$ is shorter than any leg of triangle $A'B'C'$, then to prove the lemma we shall show that another right triangle can be constructed whose sum of the angles, as in $\triangle ABC$, will be equal to two right angles and whose legs will be as large as desirable. To prove this, we join with $\triangle ABC$ an equal triangle so that its hypotenuse coincides with that of $\triangle ABC$ and such that in the resulting quadrilateral the equal sides are opposite to each other. By D we denote the vertex of the joined triangle where the corresponding angle is right (Fig. 11). Since the sum of the interior angles in each of the right triangles ABC and ABD is equal to two right angles, it is evident that the interior angles of quadrilateral $ACBD$ are right angles. Shifting $ABCD$, we can "pave" the plane by equal rectangles as shown in Fig. 11.

It is evident that the part of the plane surface shown in the figure is a rectangle. Dividing this rectangle by a diagonal, we obtain two equal right triangles with angle-sums equal to two right angles. The legs of these right triangles can obviously be made as large as desirable.*

We see that it is indeed possible to construct a right triangle the sum of whose angles is equal to two right angles and whose legs are greater than the legs of the right angled triangle $A'B'C'$. This to-

*Here Archimedes' axiom has been used (see § 4).

gether with Lemma 5 implies that the sum of the angles of (an arbitrarily chosen) right triangle $A'B'C'$ is equal to two right angles.

With the help of this lemma we can now prove the proposition mentioned earlier.

Proposition 3. *If the sum of the interior angles of at least one triangle is equal to two right angles, then so is the sum of the angles of any other triangle.*

Let two triangles ABC and $A'B'C'$ be given and assume that the sum of the angles of $\triangle ABC$ is two right angles. We shall show that the sum of the angles of $\triangle A'B'C'$ is also two right angles.

Let us construct the altitudes of the two triangles. Each of them will have at least one vertex such that the foot of the altitude from this vertex will lie inside the opposite side. Without loss of generality, we may assume that A and A' are such vertices of ABC and $A'B'C'$, respectively (this can be always achieved by relabeling the vertices).

Let P denote the foot of the altitude drawn from A in $\triangle ABC$ and P' denote that drawn from A' in $\triangle A'B'C'$. According to Lemma 4

$$D(ABP) \leq D(ABC)$$

By hypothesis, $D(ABC) = 0$, and since $D(ABP) \geq 0$ by Proposition 2, we conclude that $D(ABP) = 0$.

Thus the sum of the angles of right triangle ABP is equal to two right angles. Then, by Lemma 6, for every right triangle the sum of the angles is two right angles. But, by Lemma 3,

$$D(A'B'C') = D(A'B'P') + D(B'P'C')$$

Since $\triangle A'B'P'$ and $B'P'C'$ are right triangles, it follows from what has been proved just now that $D(A'B'P') = 0$ and $D(B'P'C') = 0$.

Hence $D(A'B'C') = 0$; consequently, the sum of the interior angles of $\triangle A'B'C'$ is equal to two right angles. ■

Having established Propositions 1-3, we can prove that there exists at least one triangle the sum of whose interior angles is equal to two right angles. Had we been really successful in proving this, then, by Proposition 3, in every triangle the sum of the interior angles would be equal to two right angles, and then, by Proposition 1, the fifth postulate would have been proved.

Here we cite one example of false proof.

Let there be given an arbitrary acute angle with vertex at O (Fig. 12). On one of its sides we take a point B and from it draw the perpendicular BA to the other side. By Proposition 2, the sum of the angles of $\triangle OAB$ does not exceed two right angles: $D(OAB) \geq 0$.

For our purpose it is enough to show that $D(OAB) > 0$ is impossible.

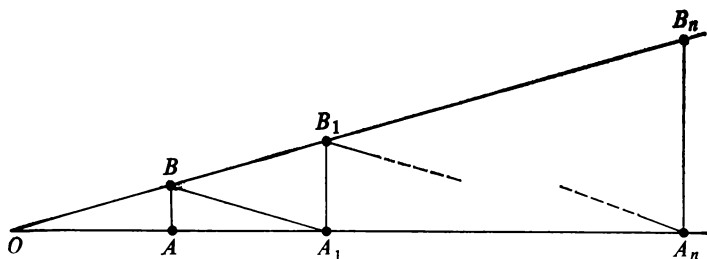


Figure 12

Assuming the contrary, we put $D(OAB) = \varepsilon > 0$. On the side OA of the given angle take a point A_1 such that $OA = AA_1$. Join B with A_1 and erect a perpendicular to OA at A_1 . Let B_1 be the point of intersection of this perpendicular with OB . According to Lemma 3,

$$D(OA_1B_1) = D(OAB) + D(BAA_1) + D(BA_1B_1)$$

But evidently the triangles OAB and BAA_1 are equal and therefore

$$D(OAB) = D(BAA_1) = \varepsilon$$

This and the previous equality yield

$$D(OA_1B_1) \geq 2\varepsilon$$

Now on the side OA of the given angle take a point A_2 such that $OA_1 = A_1A_2$. Erect a perpendicular to OA at A_2 and by B_2 denote the point of intersection of this perpendicular with OB . Similar arguments yield

$$D(OA_2B_2) \geq 4\varepsilon$$

Continuing this process, we construct a triangle OA_nB_n whose defect satisfies $D(OA_nB_n) \geq 2^n\varepsilon$. Choosing n large enough, the inequality $2^n\varepsilon > \pi$ can be satisfied. But by definition, the defect of a triangle cannot exceed π .

Thus the assumption $\varepsilon > 0$ leads to a contradiction. By the same token we have established that the defect of the triangle OAB is zero, that is, for this triangle the sum of the angles is equal to two right angles. This also establishes the fifth postulate.

It is easy to discover the weak point of the argument. Namely, the argument would have been totally justified if we had shown that the perpendiculars to OA erected at $A_1, A_2, \dots$, etc. intersect OB . However, we used the points B_1, B_2 , etc. believing them to be obvious without establishing their existence. A precise analysis shows that the existence of points B_1, B_2 , etc. cannot be proved without using the fifth postulate (this will be explained in detail later).

Thus, the above argument gives only another new equivalent of the fifth postulate. This result will be of importance for our later work; we therefore formulate it as a separate proposition.

Proposition 4. *If there is an acute angle such that the perpendicular erected at any point on one of its sides intersects the second side, then the fifth postulate holds.*

It is easy to discover the close connection between the arguments of Legendre and those of Saccheri and Lambert. In fact, the three assumptions made by Legendre regarding the sum of the angles of a triangle correspond to the obtuse, right and acute angle hypotheses of Saccheri. If we assume the obtuse angle hypothesis for a Saccheri's quadrilateral, then by dividing this quadrilateral by a diagonal we obtain two triangles in at least one of which the sum of the angles will be greater than two right angles. Conversely, if we assume that the sum of the angles of a triangle is greater than two right angles, then this leads to Saccheri's obtuse angle hypothesis. Proposition 2 thus expresses the contradictoriness of the obtuse angle hypothesis.

If we assume that the sum of the angles in a triangle is less than two right angles, then obviously for every Saccheri's quadrilateral we shall have to accept the acute angle hypothesis. Conversely, if the acute angle hypothesis is assumed to be true for any Saccheri's quadrilateral, then dividing this quadrilateral by a diagonal into two triangles we shall find that at least in one of them the sum of the angles is less than two right angles. But then, by previous propositions, in every triangle the sum of the angles will be less than two right angles, and hence the summit angles in every Saccheri's quadrilateral will be acute.

Thus we may assert

Proposition 5. *If the acute angle hypothesis is assumed to be true for one Saccheri's quadrilateral, it is also true for every Saccheri's quadrilateral.*

Finally, it is easy to see that Saccheri's right angle hypothesis and Legendre's assumption regarding the existence of a triangle in which the sum of the angles is equal to two right angles are both equivalent to the fifth postulate.

In spite of serious efforts, Legendre could not show that there is no triangle with the sum of its angles less than two right angles, which was also the case with Saccheri, who could not produce a contradiction regarding the acute angle hypothesis. But in constructing a system based on hypotheses repudiating the fifth postulate, Saccheri and Lambert went considerably farther than Legendre did.

It should be noted that Propositions 1-3 were known even before Legendre. In any case, both Saccheri and Lambert knew quite well the connection between the fifth postulate and the assertion that the sum of the angles of a triangle is equal to two right angles.

Propositions 1-3 are attributed to Legendre by tradition, since he was the first to formulate them in clear terms and since they became known due to his works.

1.3. Lobachevsky and his geometry

§ 9. Until the beginning of the nineteenth century no attempt to prove the fifth postulate was successful. In spite of the efforts of geometers over a period of more than twenty centuries, the problem of justifying the theory of parallels remained roughly at the same stage as it was in the time of Euclid.

But the first decades of the nineteenth century saw the fifth postulate problem solved; however, the mathematical community of that period did not expect such a solution and was not prepared to accept it.

The credit for solving this famous problem goes to Nikolai Ivanovich Lobachevsky (1793-1856), who from 1827 to 1846 was professor and rector at the University of Kazan. In a paper he presented before the Physico-Mathematics Faculty of that university (11 February 1826, according to the old calendar*) and in his other works**, whose publication began in 1829, Lobachevsky for the first time clearly expressed and proved the idea that the fifth postulate cannot be deduced from the other postulates of geometry. To this end Lobachevsky retains the basic premises of Euclid, except the parallel postulate, and, assuming that the parallel postulate does not hold, constructs a logical system in which the various propositions follow from accepted premises.

Many of the propositions obtained by Lobachevsky were obtained earlier by Saccheri and Lambert in connection with the acute angle hypothesis. This is understandable since the acute angle hypothesis of Saccheri is equivalent to the basic premises of Lobachevsky. But whereas the aim of Saccheri was to show that the acute angle hypothesis leads to a contradiction and hence must be rejected as being logically unacceptable, Lobachevsky, while developing a system of his theorems, establishes that this system forms a new geometry (called by him "imaginary") which, like Euclidean geometry, is free from logical contradictions.

Lobachevsky developed his "imaginary geometry" to the same level as Euclidean geometry. In this process he did not encounter any logi-

* This corresponds to 23 February 1826, according to the new calendar adopted after the Great October Revolution in 1917. Until then the calendar in Czarist Russia was twelve days behind the Gregorian calendar.

** See N. I. Lobachevsky: *Complete Works*, Gostekhizdat, Moscow-Leningrad, 1946-51 (in Russian). For a detailed account of the life and works of Lobachevsky see V. Kagan: *N. Lobachevskii and his Contribution to Science*, Foreign Languages Publishing House, Moscow, 1957.

cal contradictions. However, Lobachevsky realized this alone was not sufficient to conclude that the new geometry was really free from logical contradictions, since if there were contradictions, it would be difficult to know beforehand in which stage of the development they might be discovered. In order to prove that his geometry was free of contradictions, Lobachevsky had recourse to a deep algebraic analysis of its equations, and in this manner solved the problem as far as was possible in his time.

The proof of the consistency of Lobachevsky's geometry at the modern level of rigour was given at the end of the nineteenth century, after the general principles of logical substantiation of geometry were enunciated.

The results of Lobachevsky's investigations can be summed up as follows:

(1) The parallel postulate is not a necessary consequence of the other postulates of geometry (in other words, the former is logically independent of the remaining postulates).

(2) The fifth postulate does not follow from the other postulates because, apart from Euclidean geometry in which this postulate is true, there is another, "imaginary", geometry in which this postulate fails to be valid.

Lobachevsky was a scientist with a materialistic outlook. His materialistic views were clearly and persistently expressed in his works. He unreservedly rejected the possibilities of *apriori* knowledge; in particular, he rejected Kant's thesis that our notions of space are innate and they do not arise out of experience. In his work "On the Foundations of Geometry" (1829) Lobachevsky wrote:

The fundamental ideas with which a science emerges must be clearly formulated and must be reduced to a minimum number. Only then can they serve as a strong and sufficient base for knowledge. Such ideas are acquired by reasonings; one must not believe the innate.

Lobachevsky had a deep and subtle perception regarding the relationship between Euclidean geometry and his non-Euclidean geometry. Both geometries were logically consistent and therefore it was of no use trying to show logically that the only real geometry was the former; the question as to which one of them corresponds more closely to the properties of space must be solved by experience. In his "Imaginary Geometry" (1835) Lobachevsky wrote:

Based on astronomical observations, I proved in my works on the elements of geometry that in a triangle whose sides are almost as long as the distance between the Earth and the Sun the angle-sum can differ from

two right angles by an amount not exceeding 0.0003 of a second (the sixtieth part of a minute of angular measure). The proposition of practical geometry must therefore be viewed as having been rigorously established and, at the same time, I am convinced that independently of experience it would be unprofitable to attempt a proof of such a truth which does not automatically come under our notions about bodies.

Lobachevsky called Euclidean geometry "practical" and his own geometry "imaginary". This, however, does not mean that he considered his geometry a purely logical system (complete, we would say today). Lobachevsky recognized it as a useful instrument for mathematical analysis and in this direction wrote a comprehensive book, "The Application of Imaginary Geometry to Certain Integrals" (1836). It is interesting to note that the table of definite integrals compiled by D. Bierens de Haan (the publication of the tables was started during the lifetime of Lobachevsky in 1853 and completed in 1858) contains more than two hundred integrals that were already computed and published by Lobachevsky*. At present it is known that there is a profound connection between Lobachevsky's geometry and various other branches of mathematics as well as theoretical physics.

The ideas of Lobachevsky appeared to be paradoxical to his contemporaries and were treated ironically. Very few of them could understand and weigh his work; among them mention must be made of the German Carl Friedrich Gauss (1777-1855) and the Hungarian John Bolyai (1802-1860), who worked independently of each other as well as independently of Lobachevsky on the theory of parallels. Gauss understood the idea of the new geometry, but being content with a few elementary theorems in this direction he did not develop this enough. He did not even publish his views on the foundations of geometry, fearing that he would not be understood. John Bolyai published his works three years after the first publication by Lobachevsky (without knowing about Lobachevsky's works). In his work, Bolyai presented the same theory as Lobachevsky's but in a less developed form. Like Lobachevsky, Bolyai also did not get recognition for his work.

Only after the death of Lobachevsky did the scientific world come to realize the importance of his researches, which is immense. Before Lobachevsky, Euclidean geometry was the only one to provide a conceivable notion about space. The discovery of "imaginary geometry" or, as it is now called, non-Euclidean geometry destroyed

* For details, see the *Complete Works of Lobachevsky*, vol. 3, Gostekhizdat, Moscow-Leningrad, 1951 (in Russian) p. 413.

this viewpoint. Thus the views on geometry as science and its subject matter were broadened with far reaching consequences, which led to the modern notion of an abstract space with its numerous applications in mathematics and related fields.

In the chain of these generalizations Lobachevsky's non-Euclidean geometry proved to be the first and decisive link.

1.4. The formation of the notion of geometrical space

§ 10. We know how fruitful the Hellenic period was for mathematics. The great scientists of ancient Greece enriched mathematics by adding many new facts and developed methods for their logical systemization. After the Greeks, valuable contribution to the development of mathematics was made by the peoples of India, of countries under the Arab Khalifat and especially (from the ninth to the fifteenth centuries) by the peoples of Central Asia and Transcaucasia, who developed elements of algebra and plane trigonometry. Then in the sixteenth century a new method of solving mathematical problems in the form of alphabetical symbols was developed. This turned out to be an event of immense importance without which further achievements would have been impossible. The next two centuries, the seventeenth and especially the eighteenth, were notable for intensive development of mathematical thought and new mathematical theories were propounded. During this period differential and integral calculus were developed, and the construction of analytic geometry opened new vistas for the application of algebra and analysis in solving geometrical problems and a host of other problems in the fields of mechanics and astronomy.

However, the views on geometrical space and on those notions on which geometry was based remained essentially the same as they were during the time of Euclid and other geometers of that period. Only as a result of the remarkable achievements of the nineteenth century were attained the clarity and with it the breadth of understanding of geometry and geometrical objects which are inherent in modern mathematics and which make it different from the mathematics of old times.

Many geometrical disciplines were being actively developed in the nineteenth century. Of these we mention only the three most important: the foundations of geometry, differential geometry and projective geometry. At first the lines along which they developed were quite far from each other, but towards the end of the century they became very close and some of their parts were unified. Their synthesis gave a complete and clear elucidation of many old problems of geometry, apart from giving rise to new problems which are being developed even at present.

There are two main problems in the foundations of geometry: (1) logical development of geometry on the basis of some minimum number of statements called axioms and (2) the investigation of logical dependence between various geometrical propositions.

As already noted, these problems have their origin in the time of Euclid, whose remarkable work on the fundamentals of geometry is the first known.

The investigations carried out in connection with the proof of the fifth postulate must also be attributed to the foundations of geometry, since their aim was to prove that this postulate depended on other postulates. Having established the independence of the fifth postulate, Lobachevsky obtained the first fundamental result in this field. What is more, by constructing a geometrical system different from that of Euclid, Lobachevsky broadened the understanding of the meaning of geometry itself and hence the problem of its substantiation.

An important result in this connection was obtained later by Georg Friedrich Bernhard Riemann (1826-1866), who in his 1854 inaugural lecture "On the Hypotheses that Form the Foundation for Geometry" developed analytical principles of geometry and found, in particular, a geometrical system which was different from the systems of both Euclid and Lobachevsky. In Riemannian geometry a line is determined by two points, a plane by three, two planes intersect a line, etc., but through a given point no line can be drawn parallel to a given line. In particular, one of the theorems of this geometry states that the angle-sum of a triangle is greater than two right angles. We know that if all the statements of Euclid except the parallel postulate are retained, the last two assertions must be rejected as being contradictory (see §§ 5-8). Thus in the process of developing his system, Riemann was to change Euclidean axioms more completely than Lobachevsky did.

We find that in the middle of the nineteenth century quite important advancement was made in the foundations of geometry. But even in this period the problem of developing geometry in a strictly logical manner was not solved.

At the end of the 1860s, when the ideas of Lobachevsky were ultimately accepted, the problem of the logical construction of geometry became the problem of the day. In particular, its solution was necessary in order to make the achievements of Lobachevsky completely clear. Indeed, not only did the main result of Lobachevsky (namely, that the fifth postulate is independent of the remaining postulates of geometry) fail to be rigorously proved but even the problem could not be precisely formulated unless all the geometrical postulates were known.

Towards the end of the nineteenth century many works authored by first rate mathematicians were published on this question. Of

these especially popular became David Hilbert's (1862-1943) *Foundations of Geometry*, which was published in 1899 and which won the International Lobachevsky Prize in 1903.

In his book, Hilbert formulates a complete system of axioms of Euclidean geometry, that is, a list of basic propositions from which the whole subject matter of this geometry can be obtained in a logical sequence.* Hilbert also establishes the independence of the most important axioms of his system from its remaining axioms.

Hilbert's axioms and an analysis of their relationship are discussed in the following chapters. Here we shall only touch upon that point of view from which the basic geometrical notions and axioms are examined at the present time.

In contrast to Euclid's *Elements*, the modern list of axioms of Euclidean geometry does not contain any description of geometrical figures. It is only assumed that there exist three groups of objects called "points", "lines" and "planes" in relation to which some fully defined conditions are observed as follows:

(1) Between the objects called points, lines and planes as well as between certain sets of these objects (segments, angles) there must exist relationships which are expressed by words such as "lies on...", "between", "are congruent".

(2) These relationships must satisfy the conditions stated in the axioms that follow.

Of course, the axioms are formulated having in mind the empiric subject matter accumulated in geometry and with the intention that all this material could be deduced from these axioms by logical arguments. But the objects which figure in the axioms do not necessarily have any special nature or, say, any definite external form. The relationships between these objects are also not necessarily of any special character. Both the objects and the relationships between them can be chosen in an arbitrary manner, but they must satisfy the requirements of the axioms. There are two reasons for such an attitude towards geometry and geometrical objects.

(a) Geometry uses ideas arising out of experience as a result of an abstraction from the real world, where only some properties of the real objects are taken into account; in proving any theorem by strictly logical arguments one deals only with these properties of the objects. Notice must be taken of these properties in formulating the axioms and definitions. All the remaining properties which we are accustomed to imagine when we hear the words "point", "line" and "plane" do not play any role whatsoever in a logical development of geometry and hence they must not be mentioned in the basic statements of geometry.

* Lists of the axioms of Euclidean geometry were formulated by other authors before Hilbert, for example, by Moritz Pasch (in 1882). But the list given by Hilbert turned out to be the simplest of all those known at that time.

(b) Apart from Euclidean geometry whose theorems correspond to our usual notions about the properties of geometrical figures, there exist various other geometrical systems (of Lobachevsky and Riemann) which contradict the usual notion of space. Therefore, in a sufficiently general formulation of the problem of substantiating geometry, the understanding of geometrical objects itself must be so general as to cover all the necessary cases.

According to the above exposition, we may say that the geometrical space defined by a given system of axioms is a set of objects, called geometrical elements, whose mutual relationships satisfy the requirements of the axioms of the given system.

Thus one may speak of Euclidean space meaning thereby a set of elements which satisfy the requirements of the axioms of Euclidean geometry, or of Lobachevskian space as a set of elements satisfying the requirements of the axioms of Lobachevskian geometry.

But Euclidean space itself, for example, may have many different forms depending upon what concrete objects are taken as its elements. For instance, apart from the usual notion about points, lines and planes, we may agree to call any sphere of a given radius d a "point", any infinite circular cylinder of radius d a "line", and the space between any two parallel planes which are at a distance d a "plane". The basic relationships between these objects can be defined as follows. We shall say that a "point", represented by a sphere A , lies on the "line", represented by a circular cylinder a , if the sphere is inscribed in the cylinder. We shall say that a "point" (represented by the sphere A) lies on the "plane", represented by a spatial layer α if the sphere touches the two ordinary parallel planes which bound α . Similarly, we shall say that a "point" B lies on "line" a between "points" A and C if the centre of the sphere representing B lies between the centres of the spheres representing A and C . We shall say that a figure M is equal to, or congruent to, another figure N if M can be superposed on N with the aid of some motion (in the present context M and N are assumed to be made up of "points", "lines" and "planes"). These relationships between the objects under consideration meet the requirements of the axioms of Euclidean geometry. Therefore any theorem which is deduced logically from these axioms expresses a certain fact relating to the above "points", "lines" and "planes". The set of "points", "lines" and "planes" with these relationships between them is one of the concrete forms of Euclidean space.

Choosing other objects as points, lines and planes and defining their mutual relationships so as to satisfy the requirements of the axioms of Euclidean geometry, we obtain other concrete forms of Euclidean space. To every concrete form of Euclidean space there corresponds a definite interpretation of Euclidean theorems. Naturally, Lobachevskian geometry also admits of various concrete inter-

pretations, as does any other system based on axioms (see §§ 49-61, 67, 168-171).

Thus removing from geometry all references to visual concepts and retaining only the logical framework, we have the opportunity of filling it with various concrete material. Consequently, in an abstract logical development of geometry the real base is not at all lost; on the other hand, the field in which geometry can be applied is widened.

The following remark is of great importance. The broad view of geometrical elements and geometrical axioms which has been taken above enables us to select the system of axioms itself with a degree of arbitrariness, which can be adapted to suit the particular field of investigation. In this way the axiomatic method of geometry is applied in other fields of mathematics, in mechanics, in physics, and this leads to the modern abstract spaces whose elements are sets, functions, transformations, etc. As an illustration of the application of general geometrical ideas, one may point out the Minkowski space, which plays an important role in the special relativity theory.

The foundation for the idea of an abstract space was built by the evolution of mathematics in the nineteenth century. In the problems related to the foundations of geometry, this idea was a direct outgrowth of the discovery made by Lobachevsky. But this discovery also greatly influenced the development of geometrical ideas through other branches of knowledge.

The modern views on geometrical space were formed largely by the development of differential geometry. In his memoir "General Investigations of Curved Surfaces" (1827), Gauss noted some special properties of surfaces forming its intrinsic geometry. These are the properties which can be observed by means of measurements on the surface itself (the practical source of ideas for the intrinsic geometry of surfaces came from land-surveying).

In 1868 the work of Eugenio Beltrami (1835-1900), "The Interpretation of Non-Euclidean Geometry", appeared, in which the author demonstrated that plane Lobachevskian geometry can be regarded, under certain conditions, as an intrinsic geometry of some surface. This enabled him to include non-Euclidean plane geometry together with Euclidean plane geometry in a perfectly real field of the theory of surfaces.

The common points of the axiomatic investigations of Lobachevsky with the differential-geometry methods of Gauss were to a large extent responsible, at least in two-dimensional cases, for the generalization of geometrical concepts. But the level of mathematics at that time was so high that the application of differential-geometry methods to non-Euclidean geometry could not be confined only to two-dimensional cases. Already in 1854 Riemann, in his above mentioned work "On the Hypotheses that Form the Foundation for Geomet-

ry", defined spaces which were generalizations of Euclidean and Lobachevskian spaces as well as of the space mentioned earlier in connection with Riemannian geometry. These general Riemannian spaces differ in their properties from Euclidean spaces as any curved surface differs from a plane.

The pure analytical method employed by Riemann in the investigation of geometrical problems enabled him to generalize the concept of curvature directly to multidimensional cases. The general Riemannian spaces turned out to be useful for theoretical physics and to this day are carefully studied.

Approximately at the same time when Lobachevsky was starting his investigations on the theory of parallels and Gauss was working on his theory of surfaces, there sprang up a new branch of geometry, projective geometry. Governed by pictorial concepts, projective geometry was at first thought to be quite far from the complicated axiomatic problems. But in the 1870s Felix Klein (1849-1925) gave a general interpretation of the geometrical systems of Euclid, Lobachevsky and Riemann, which was based on projective geometry. (For this Klein used some of the results obtained earlier by Arthur Cayley (1821-1859).) The studies of Klein were closely related with his conception of geometry as the theory of invariants of a certain group of transformations. The group-theoretic approach to geometry laid down by Klein in his lectures "A Comparative Review of Recent Researches in Geometry" (the views expressed in it are known as the *Erlangen Programm* (1872)) enabled him to classify the important geometrical systems and transformations connected with them.

The subject matter of the present book is divided in accordance with the three directions mentioned above along which the concept of geometrical space developed in the nineteenth century. Chapters 2, 3 and 4 are devoted to the questions of axiomatics. Projective geometry and the classification of geometrical systems according to the theory groups are dealt with in Chapters 5 and 6. A part of Chapter 7 devoted to Minkowski space also overlaps them. And, finally, in Chapters 8 and 9 geometrical systems are investigated by the methods of differential geometry.

Chapter 2

The Axioms of Elementary Geometry

2.1. Geometrical elements

In this chapter we set forth the axioms formulated by Hilbert.* Apart from the axioms, the main theorems are also given so as to make clear the general principles of the logical development of geometry.

§ 11. Three different sets of objects are considered below. The objects of the first set are called *points*, those of the second *lines*, while those of the third *planes*. The set of all points, lines and planes is called *space*.

The points, lines and planes are related to each other in ways expressed by the words "lie", "between" and "congruent". These relationships must satisfy the following axioms; apart from this, the nature of the objects and their relationships are totally arbitrary.

All the axioms are divided into five groups.**

Group I contains eight axioms of incidence.

Group II contains four axioms of betweenness.

Group III contains five axioms of congruence.

Group IV contains two axioms of continuity.

Group V contains one axiom of parallelism.

2.2. Group I: axioms of incidence

§ 12. Let us assume that lines and planes are in some relationship with points. If the line a is related to the point A , we shall say that " a passes through A ", " A lies on a ", " A is a point on a ", " A belongs to a ", " a belongs to A ". If several lines are related to A , we shall say that these "intersect at A " or "have a common point A ". If two points A and B are related to a line a , we say that " a joins A and B " or " a passes through A and B ", etc. The conditions of these relationships are expressed by Axioms I.1-I.8.

* Hilbert's axioms given here are based on the seventh edition of D. Hilbert: *Grundlagen der Geometrie*, B. G. Teubner, Berlin, 1930 (*Foundations of Geometry*, 2nd ed., Open Court, La Salle, Ill., 1971 (in English)).

** In numbering these groups we somewhat deviate from Hilbert, who put the axiom of parallels in the fourth group and the axioms of continuity in the fifth.

(I.1) *For any two distinct points A and B there exists a line a that passes through both A and B .*

(I.2) *For any two distinct points A and B there is not more than one line passing through both points.*

These two axioms can be expressed as follows: any two distinct points determine one and only one line through these points.

(I.3) *At least two points lie on every line. There are at least three points that do not lie on one line.*

Regarding a point A and a plane α which are related to each other, we shall use expressions " A lies in α ", " A is a point of α ", " α passes through A ", etc.

(I.4) *Through any three noncollinear points, A , B and C there passes a plane α . On every plane there lies at least one point.*

(I.5) *Through any three noncollinear points A , B and C there passes not more than one plane.*

(I.6) *If two points A and B of the line a lie in α , every point of a lies in α .*

In this case we say that " a lies in α ", " α passes through a ", etc.

(I.7) *If two planes α and β have a common point A , they have at least one more common point B .*

(I.8) *There are at least four points that do not lie in one plane.*

In the axioms of incidence there appear known relationships between elements of geometry which are expressed by saying "a point lies on a line", "a plane passes through a point", etc. Visual evidence for these expressions are in no sense given; only definite features necessary for the derivation of theorems are mentioned in Axioms I.1-I.8.

Conditions of Axioms I.1 and I.2 were expressed by Euclid in his first postulate and his ninth axiom. But Euclid could hardly feel the necessity of Axiom I.3 and of the other axioms.

Understandably, a geometer who leaves even the slightest possibility of using spatial intuition in his arguments will not postulate that there exist two points on a line or that there are three points not lying on a line, etc. Visual evidence would immediately lead him to demand of the axiom that there exist infinitely many points on a line. However, this condition need not be formulated as an axiom since it will be proved below. This explains the desire for keeping the requirements of the axioms to a minimum.

Using Axioms I.1-I.8, certain theorems can already be proved. We cite some of them below.

Theorem 1. *Two lines have at most one common point. Two planes have either no common points or have a common line on which all the*

common points of these two planes lie. A plane and a line not lying in it have at most one common point.

The first assertion of this theorem follows from Axiom I.2.

Proof of the second assertion. Let two planes α and β have a common point A . By Axiom I.7, they have one more common point B . According to Axiom I.6, a line a joining A and B consists of common points of α and β . More than that, a comprises all the common points of α and β . Indeed, suppose that α and β have another common point C that does not lie on a . Then Axiom I.5 implies that α and β are not distinct, since they have three common points not lying on one line.

The third assertion is a consequence of Axiom I.6. ■

Theorem 2. *One and only one plane passes through a line and a point not lying on this line or through two lines with a common point.*

Proof. Take a line a and a point A not on it. By Axiom I.3, there are two points B and C on a . By the hypothesis and Axiom I.2, the points A, B, C are noncollinear. According to Axiom I.4, there exists a plane α passing through A, B and C ; by Axiom I.6, α passes through a . There is no other plane passing through A and a . Indeed, if there were another plane α' through A and a , then two distinct planes α and α' would pass through A, B and C , violating Axiom I.5. ■

Theorem 3. *Every plane contains at least three points.*

Proof. Let a plane α be given. By Axiom I.4, α contains a point A , and, by Axiom I.8, there is a point B not lying in α . According to Axiom I.3, there is another point C that does not lie on the line AB . The planes ABC and α have a common point A , and Axiom I.7 states that these planes have another common point D . Thus α necessarily contains, besides A , another point D . By Axiom I.8, there is a point E that does not lie in ABD , while Axiom I.4 asserts the existence of a plane ABE distinct from ABD . Applying Axiom I.7, again we conclude that the planes ABE and α have a common point F (which does not lie on AB , in view of Axiom I.6). Since D and F do not lie on AB , the second assertion of Theorem 1 implies that these points cannot be common points of the planes ABD and ABF . Hence D and F are distinct. Thus there are three points A, D and F in α . ■

We carried out the proofs in detail so as to give the reader an idea how elementary geometry can be logically developed using the accepted axioms. Any appeal to diagram or visual evidence has been completely ruled out in the above discussions; every assertion is established by using the axioms or the theorems previously proved.

It should be noted that only a few geometrical facts can be established by means of Axioms I.1-I.8. In particular, these axioms do not enable us to show that the set of geometrical elements is infinite (for details, see § 70).

2.3. Group II: axioms of betweenness

§ 13. We assume that a point on a line is related to two other points on the same line and this relationship will be expressed by saying "lies between". This relationship must satisfy the following axioms.

(II.1) *If a point B lies between a point A and a point C , then A , B and C are three distinct points of a line, and B also lies between C and A .*

(II.2) *For any points A and C there exists at least one point B on the line AC such that C lies between A and B .*

(II.3) *Among any three points of a line there is only one point lying between the other two.*

Axioms II.1-II.3 are known as *linear axioms of betweenness*.

Definition 1. The pair of points A and B is called a *segment* and is denoted by AB or BA . Points lying between A and B are called *interior points*, or simply *points of the segment AB* , while the points A and B are called the *endpoints* of the segment. All other points of the line AB are called *exterior to the segment AB* .

Remark. Axioms II.1-II.3 do not assert that there are other points between A and B . Therefore from these axioms it is not clear if every segment contains interior points. However, Axiom II.2 states that every segment has exterior points.

Apart from linear axioms of betweenness II.1-II.3, Group II contains one more axiom concerning the location of geometrical elements on the plane.

(II.4—Pasch's axiom) *Let A , B , C be three noncollinear points, and let a be a line in the plane ABC not containing any of the points A , B , C . Then if a passes through a point of segment AB , it also passes through either a point of segment AC or a point of segment BC .*

2.4. Corollaries of the axioms of incidence and betweenness

§ 14. By using the axioms of incidence and betweenness, we can prove many important facts of geometry. First we shall establish two theorems supplementing Axioms II.1-II.3.

Theorem 4. *Given any two points A and C , there exists at least one point D on the line AC which lies between A and C .*

Proof. By Axiom I.3, there exists a point E lying outside the line AC , and, by Axiom II.2, there is a point F on AE such that E is a point of the segment AF (Fig. 13). Again by Axiom II.2, there is a point G on FC such that C lies between F and G . Then Axiom II.3 implies that G is not between F and C , that is, is not on the segment FC . According to Pasch's axiom (Axiom II.4), EG must intersect either the segment AC or the segment FC . But the line EG cannot intersect the segment FC , for then Axioms I.1 and I.2 of

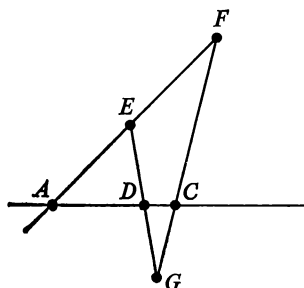


Figure 13

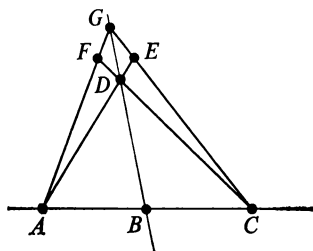


Figure 14

incidence would imply that all the points in question lie on one line, which is not the case. Consequently, the line EG intersects the segment AC at some point D , which establishes the existence of a point D between A and C . ■

Theorem 5. *Among any three points A, B, C on a line there always exists one point lying between the other two.*

Proof. Suppose that A does not lie between B and C , and C does not lie between A and B . By Axiom I.3, there is a point D that does not lie on the line AC . By a line join this point with B (Fig. 14). By Axiom II.2, there exists a point G on BD such that D lies between B and G . Applying Axiom II.4 to $\triangle BCG$ and the line AD , we find that AD and CG intersect at a point E that lies between C and G . Similarly, we can show that CD and AG intersect at some point F lying between A and G . Again applying Axiom II.4 to the triangle AEG and the line CF , we see that D lies between A and E , while another application of this axiom to the triangle AEC and the line BG shows finally that B lies between A and C (since G does not lie between E and C). ■

Axioms II.2 in conjunction with Theorem 4 and Axiom II.3 in conjunction with Theorem 5 yield the following two theorems.

Theorem A. *Given any two points A and C , there are points lying inside the segment AC and points lying outside this segment on the line AC .*

Theorem B. *Of three points of a line, one and only one always lies between the other two.*

Now we give an important supplement to Pasch's axiom in the form of the following theorem.

Theorem 5a. *If the points A, B and C are noncollinear and if a line a intersects any two of the three segments AB, BC, AC , then it does not intersect the third.**

* By a line intersecting a segment we mean that the former contains some interior point of the latter.

We prove this theorem by contradiction. We shall show that the assumption that a intersects each of the segments AB , BC and AC at points P , Q , R , respectively, leads to a contradiction. ■

First, it is clear that B does not lie on the line PQ (otherwise all the points A , B , C would lie on PQ). Next, we find that R lies outside the segment PQ , since otherwise the line AC intersecting the side PQ of the triangle PQB must intersect, by Pasch's axiom, the other side BQ also, that is, C must lie between B and Q , contradicting the hypothesis (according to the assumption, Q lies between B and C , and since of three points only one can lie between the remaining two, this rules out the possibility of C being between B and Q). Similarly, we can show that P lies outside the segment QR and Q lies outside the segment PR . This contradicts Theorem B. ■

Later we shall need the following two lemmas.

Lemma 1. *If B lies on the segment AC and C lies on the segment BD , then B and C lie on the segment AD .*

Proof. In view of Axioms I.3 and II.2, we can choose a point E not on the line AB and a point F on EC such that E lies between C and F (Fig. 15). The point B lies on the segment AC . Therefore applying Axiom II.4 to the triangle AEC and the line FB , we conclude that the line FB must intersect either segment AE or EC . Since E lies between F and C , Axiom II.3 implies that F cannot lie between E and C . Therefore FB must intersect the segment AE . Applying Axiom II.4 to the triangle FBC and the line AE and again using Axiom II.3, we find that the point of intersection of the segment AE and the line FB lies between F and B . Denote this point of intersection by G . It can be similarly shown (by applying Axiom II.4 to the triangle GBD and the line CF and then using Axiom II.3) that CF intersects GD at a point H . Since H lies on the segment GD , and E , by Axiom II.3, does not lie on the segment AG , Axiom II.4 implies that the line EH has a common point with the segment AD , that is, C lies on the segment AD . The proof that B also lies on AD is exactly the same. ■

Lemma 2. *If C lies on the segment AD and B on the segment AC , then B also lies on the segment AD and C on the segment BD .*

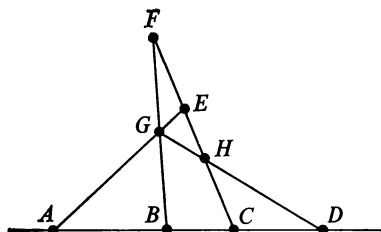


Figure 15

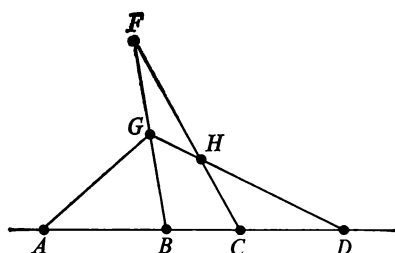


Figure 16

Proof. Outside the line AB take a point G and then choose a point F such that G lies on the segment BF (Fig. 16). Axioms I.2 and II.3 imply that the line CF does not have any common points either with the segment AB or with the segment BG . Then, according to Axiom II.4, it does not have any common points with segment AG either. Noting that C lies on the segment AD and applying Axiom II.4 to the triangle AGD , we find that CF must intersect GD at some point H . This result and another application of Axiom II.4 to the triangle BGD imply that the line FH intersects the segment BD . Thus C lies on the segment BD . The first assertion of Lemma 2 follows now from Lemma 1. ■

The following important theorem can now be easily proved.

Theorem 6. *Between any two points of a line there are infinitely many other points of this line.*

Proof. Let A and B be two points of a line a . By Theorem 4, there is a point C between A and B and a point D between A and C . By Lemma 2, D also lies between A and B ; accordingly, A, B, C and D are distinct points of a . Similarly, we can show that between A and D there lies a point E and that this point also lies between A and C and between A and B . Therefore, A, B, C, D and E are distinct points. Continuing in this manner, it can be established that between A and B there are infinitely many points $C, D, E, \dots$ ■

Note that Lemmas 1 and 2 imply the following proposition.

Suppose that each of the points C and D lies between the points A and B . Then, if M lies between C and D , it also lies between A and B .

Indeed, by Theorem B (p. 46), of the three points A, C, D one and only one can lie between the other two. But A cannot lie between C and D since this would contradict Lemma 1. To be definite, assume that C lies between A and D (if necessary, we can redesignate C and D). Then the position of the points D, M, C, A satisfies the same conditions as do the positions of A, B, C, D in Lemma 2. Therefore, by this lemma, M lies between A and D . Now we can assert that the position of A, M, D, B also satisfies the same conditions as the position of the points A, B, C, D in Lemma 2. As a consequence of this lemma, M lies between A and B . ■

Thus we have proved the following

Theorem 7. *If the points C and D lie between the points A and B , all the points of the segment CD belong to the segment AB .*

Definition 2. In this case it is said that *segment CD lies inside segment AB* .

As a corollary to Lemma 2, we at once have

Theorem 8. *If a point C lies between the points A and B , all the points of the segment AC belong to the segment AB .*

Similarly, from Lemma 2 (using Axiom II.3) one can prove (by contradiction)

Theorem 8a. *If a point C lies between the points A and B , no point of the segment AC can be a point of the segment CB .*

Somewhat complicated is the proof of

Theorem 8b. *If a point C lies between the points A and B , each point of the segment AB distinct from C belongs either to the segment AC or to the segment CB .*

Proof. Let M be a point of AB distinct from C . Assume that M does not belong to either AC or CB . Then either C lies between A and M or A lies between C and M . If C lies between A and M , then, since M lies between A and B , the second assertion of Lemma 2 implies that M lies between C and B , contrary to the assumption. If A lies between C and M , then, since C lies between A and B , Lemma 1 implies that A lies between M and B , accordingly, M cannot lie between A and B . This again contradicts the assumption. ■

On the basis of Theorems 8, 8a, 8b, one can assert that the set of interior points of the segment AB , excluding the point C , is the sum of the sets of the interior points of AC and CB , and these two segments do not have any common point.

Definition 3. Let O be a given point of a line a and let A and B be two other distinct points of the same line. If O does not lie between A and B , we shall say that A and B are situated on a on the same side of O . If O lies between A and B , we shall say that A and B are situated on a on opposite sides of O .

Theorem 9. *A point O of a line a divides all the other points of this line into two nonempty classes such that any two points of a belonging to the same class lie on the same side of O , while two points belonging to different classes lie on opposite sides of O .*

In order to prove this result, take a point A distinct from O on a and form one class containing points for which A lies on the same side of O and another class containing points for which A lies on the opposite side of O . Then it has to be proved that (i) each class is nonempty, (ii) each point, except O , of a belongs to only one class, (iii) if M and N belong to the same class, the point O does not belong to the segment MN , and (iv) if M and N belong to different classes, the point O belongs to the segment MN .

All these statements can be easily proved by Theorems 8, 8a and 8b.

Definition 4. A point O of a line a together with another point A of this line is said to define a *half line*, or *ray*, OA ; the points lying with A on the same side of O are called *points of the half line* OA , and O is called the *vertex of the half line* OA .

If A' is a point of the half line OA , then the half lines OA and OA' are identical in the sense that each point of the half line OA' is also a point of the half line OA and conversely.

Theorem 9 implies that any point O on a defines exactly two half lines with a common vertex at O .

The above exposition enables us to treat the set of points of every line as a set ordered in a definite manner. As is known, a set is *ordered* if the two notions "precedes" and "follows" are defined on this set such that of its two distinct elements x and y the first precedes the second; in this case the second is also said to follow the first. The transitivity condition must be satisfied: if x, y, z are three elements and x precedes y and y precedes z , then x precedes z .

The set of real numbers can be ordered according to "magnitude", saying that a precedes b if and only if $a < b$.

Let a be any line and O a point on a . Consider one of the two half lines with the common vertex at O . We shall say that a point A precedes point B on this half line if A belongs to the segment OB .

Lemma 2 immediately implies that if A precedes B and B precedes C , then A precedes C . Hence the points of every half line are seen to be in a definite order.

Let us now agree to call one of the two half lines with the common vertex O the *first* half line and define the order of points on the entire line a in the following manner.

(1) Let A, B be two points of the first half line. Then A precedes B on a if B precedes A on the first half line.

(2) All the points of the first half line precede on a the point O .

(3) All the points on the first half line precede on a the points of the second half line.

(4) Point O precedes on a the points of the second half line.

(5) Let A and B be two points of the second half line. Then A precedes B on a if A precedes B on the second half line.

For any two points of a , Conditions 1-5 determine which point precedes the other.

The transitivity condition is fulfilled.

Indeed, let three points A, B, C be given on a , and suppose that A precedes B and B precedes C in the sense of Conditions 1-5. We shall show that under the same conditions A precedes C .

If all three points lie on one of the two half lines with vertex O , then, as noted above, Lemma 2 ensures that A precedes C .

If A lies on the first half line and B on the second (or coincides with O), then C is definitely a point of the second half line (otherwise there would be a contradiction with Condition 3 or 2). In this case A precedes C , by Condition 3.

If the points A and B lie on the first half line and C lies on the second or coincides with O , then A precedes C , by Condition 3 or 2.

All other possible locations of A, B, C would contradict Conditions 1-5.

This shows that the transitivity property holds.

If we interchange the first and second half lines and again impose Conditions 1-5, we shall obtain a new order for the points of a which will be the *reverse* of the original order, that is, if A precedes B in the first ordering, then B precedes A in the second.

Let O' be a point of a distinct from O . Choosing one of the two half lines with common vertex O' to be the first, we can again order the points of a using Conditions 1-5. This order will coincide with one of the two orders obtained earlier by choosing O as the vertex (the proof is omitted). Thus, irrespective of the choice of O , Conditions 1-5 completely define two possible orders, the reverse of each other, for the points of a .

Choosing one of these orders, we shall say that a *direction* is defined on the line.

From the definition of the ordering of points on a line, the following proposition can be demonstrated: *If B lies between A and C , then either A precedes B and B precedes C , or C precedes B and B precedes A . Conversely, if A precedes B and B precedes C , or C precedes B and B precedes A , then B lies between A and C .*

In other words, points on a line are ordered in such a manner that the position of B between A and C in the sense of the present ordering is equivalent to the position of B between A and C in the sense of the ordering of § 13.

§ 15. The foregoing discussions were concerned with the position of points on a line. We shall now establish a number of propositions characterizing the location of points in the plane and in the space.

Theorem 10. *Any line a lying in a plane α divides the points of this plane, not lying on a , into two nonempty classes such that any two points A and B of different classes determine a segment AB containing a point of a , while any two points A and A' of the same class determine a segment AA' which does not contain any point of a .*

Proof. Take any point P in α not on a , and to the first class assign every point A of the plane, not lying on a , such that the segment PA does not contain points of a , the point P itself is also assigned to this class (Fig. 17). To the second class assign all points B not on a such that the segment PB contains a point of a . Then

(1) each class is nonempty. Indeed, if Q is some point of a , then, by Axiom II.2, there is a point R on PQ such that Q lies between P and R , accordingly, R belongs to the second class. The first class, on the other hand, contains, for instance, P ;

(2) all points of α (except those lying on a) belong to only one class. In fact, any segment either contains or does not contain a point of a .

(3) any two points A and A' of the first class determine a segment AA' that does not contain any point of a .

Indeed, assuming that AA' contains a point of a and that P , A , A' are noncollinear we see that Pasch's axiom (Axiom II.4)

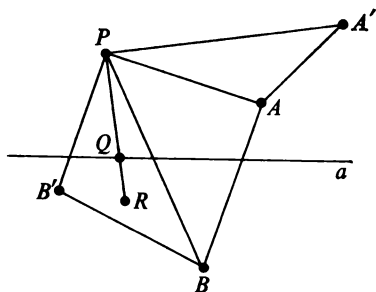


Figure 17

implies that one of the two segments PA or PA' must contain a point of a , which contradicts the hypothesis. If P, A, A' are collinear, the same conclusion is drawn by applying Theorem B and Theorem 8 when P does not belong to the segment AA' or by applying Theorem 8b when P belongs to the segment AA' .

(4) Any two points B and B' of the second class determine a segment BB' that does not contain any point of a .

This is a consequence of Theorem 5a when P, B, B' are noncollinear and of Theorem B and Theorem 8a when P, B, B' are collinear.

(5) Any two points A and B of different classes determine a segment AB which contains some point of a .

Indeed, according to the condition, the segment PB contains a point of a . If P, A, B are noncollinear, then, by Pasch's axiom, either PA or AB contains a point of a , but the given condition rules out this possibility for PA . Therefore AB contains a point of a .

If P, A, B are collinear, the same conclusion is drawn by applying Theorem B and Theorems 8 and 8b. ■

Remark. It is easy to show the following. (1) Each class contains infinitely many points (Theorem 6 is to be applied for the proof). (2) Taking P' as any point of the first class and dividing the points of the plane again into two classes as above but replacing P by P' , we obtain the same classes as above. (3) If P is replaced by any other point of the second class, this will affect only the labelling of the classes.

Definition 5. In the notation of Theorem 10, we shall say that the points A and A' lie in α on the same side of a , while A and B lie in α on the opposite sides of a .

Theorem 11. Any plane α divides the points of the space, not lying in α , into two nonempty classes such that any two points A and B from different classes determine a segment AB which contains a point of α , while any two points A and A' from the same class determine a segment AA' that does not contain any point of α .

Definition 6. Points A and A' are said to lie in space on the same side of α , while the points A and B lie on the opposite sides of α .

We omit the proof of Theorem 11, but remark that although this theorem concerns space geometry, the proof does not require any new axioms of order, apart from Axioms II.1-II.4 which were introduced for points on a line and a plane.

The axioms of Group II define the important notions regarding the order of points on a line, the location "on the same side or opposite sides", etc. Of these the notion "lies between" is fundamental, all the remaining notions being based on this.

With the aid of Axioms II.1-II.4 one can define in a natural way a polygonal line, a triangle, a polygon in general, and can prove the proposition regarding the division of the plane by a simple polygon into two domains. But these axioms do not imply, in particular, that the set of elements of geometry is uncountable (for this, see Chapter 4, § 72).

2.5. Group III: axioms of congruence

§ 16. We assume that a segment is related to another segment (or to itself), and this relation will be expressed by the word "is congruent" or "is equal". The congruence relation must satisfy the following axioms.

(III.1) *If A and B are two points on a line a and A' is a point on the same line or on another line a' , then one and only one point B' can be found on a' on a given side of A' such that the segment AB is congruent to the segment $A'B'$.*

This relationship between AB and $A'B'$ is expressed by

$$AB \equiv A'B'$$

For any segment AB , it is required that

$$AB \equiv BA$$

The first part of this axiom can be expressed in short as follows: *Every segment can be uniquely superposed on any given side of a given point on any given line* (Fig. 18).

(III.2) *If the segments $A'B'$ and $A''B''$ are both congruent to the segment AB , then $A'B'$ is congruent to $A''B''$, that is, if*

$$A'B' \equiv AB \quad \text{and} \quad A''B'' \equiv AB$$

then

$$A'B' \equiv A''B''$$

Axioms III.1 and III.2 imply that if $AB \equiv A'B'$, then $AB \equiv B'A'$. Indeed, the two relations

$$AB \equiv A'B', \quad B'A' \equiv A'B'$$

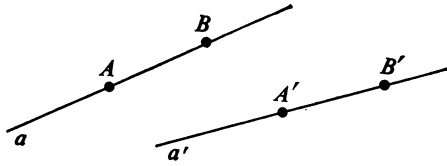


Figure 18

(the second is a consequence of Axiom III.1) and Axiom III.2 imply that $AB \equiv B'A'$. This together with Axiom III.1 yields

Corollary. *Each segment is congruent to itself, that is,*

$$AB \equiv AB, \quad AB \equiv BA$$

Indeed, the relation $AB \equiv BA$ is ensured by Axiom III.1, while $AB \equiv AB$ follows from $AB \equiv BA$ on the basis of the previous result.

Next, we prove that: *if $AB \equiv A'B'$, then $A'B' \equiv AB$, that is, the congruence relation between the segments is symmetrical.*

Indeed, we have $A'B' \equiv A'B'$. If $AB \equiv A'B'$ is given, then these two relations together with Axiom III.2 imply that $A'B' \equiv AB$.

Finally, let us show that if

$$AB \equiv A'B' \quad \text{and} \quad A'B' \equiv A''B''$$

then also

$$AB \equiv A''B''$$

that is, *the congruence relation between the segments is transitive.*

For the proof, note that by the previous result the relations $AB \equiv A'B'$ and $A'B' \equiv A''B''$ yield

$$AB \equiv A'B' \quad \text{and} \quad A''B'' \equiv A'B'$$

Now $AB \equiv A''B''$ is ensured by Axiom III.2.

Thus Axioms III.1 and III.2 enable us to establish the following: (1) each segment is congruent to itself, (2) the congruence of segments is independent of the order of their points*, and (3) the congruence relation of the segments is symmetric and transitive.

For more serious results we shall require new axioms.

(III.3) *Let AB and BC be two segments on a line a having no common interior point and let $A'B'$ and $B'C'$ be two segments on the same*

* This means that $AB \equiv A'B'$ implies that $AB \equiv B'A'$, $BA \equiv A'B'$ and $BA \equiv B'A'$. The first has been demonstrated above and the last two can be shown using the symmetry and transitivity of the congruence relations of segments.

line or on another line a' , again without any common point. If, in addition,

$$AB \equiv A'B' \quad \text{and} \quad BC \equiv B'C'$$

then (Fig. 19)

$$AC \equiv A'C'$$

Definition 7. A pair of half lines h and k which originate from the same point O and do not lie on one line is called an *angle*. To denote an angle, we shall use $\angle(h, k)$ and $\angle(k, h)$.

For two points A and B on h and k the angle will also be denoted by $\angle AOB$.

The half lines h and k are called *sides* of the angle and O its *vertex*.

Let h' be a half line which with h completes a line and k' a half line which with k completes a line. A point of the plane that lies on the same side of the line h , h' as the points of the half line k and on the same side of the line k , k' as the points of the half line h is called an *interior point* of $\angle(h, k)$ and the aggregate of all such interior points is called the *interior of the angle*. All other points of the plane, except O and the points of the half lines h and k , containing the angle, are called *exterior points* of the angle and their aggregate forms the *exterior of the angle*. (In Fig. 20 the interior domain of $\angle(h, k)$ is shown by double shading.)

We note the following theorem.

Theorem 11a. If A and B are points on different sides of an angle, then any half line through the vertex inside the angle intersects the segment AB . Conversely, any half line joining the vertex with a point of AB lies in the interior of the angle.

Proof of the first part. Let $\angle(h, k)$ be the given angle (A lies on h), and let l be a half line originating from the vertex in the interior. On the half line h' completing h take any point C and consider the triangle ABC . Let l' denote the half line completing l and l^* the line composed of l and l' . By Axiom II.4, l^* must intersect either CB or AB . But l^* does not have any point inside $\angle(h', k)$; accordingly, it must intersect AB .

Furthermore, the half line l' does not have any point inside $\angle(h, k)$; therefore the segment AB intersects l . This establishes the first part of the theorem.

For the proof of the second part, only trivial arguments are required. ■

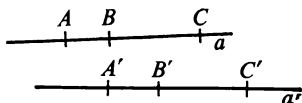


Figure 19

Now we shall introduce the last important notion, that of congruence of angles. We assume that an angle is related to another angle (or to itself) and express the relation by "is congruent" or "is equal".

(III.4) Let an angle $\angle (h, k)$ be given in a plane α , and let a' be any line in this plane or in any other plane α' ; suppose that a definite side of α' with respect to a' is assigned.

Denote by h' some ray of a' with the vertex at O' . Then there is one and only one ray k' in α' such that $\angle (h, k)$ is congruent to $\angle (h', k')$; moreover, all the interior points of $\angle (h', k')$ lie on a given side of a' . This congruence will be denoted by

$$\angle (h, k) \equiv \angle (h', k')$$

If $\angle (h, k) \equiv \angle (h', k')$, then $\angle (k, h) \equiv \angle (k', h')$. Every angle is congruent to itself, that is,

$$\angle (h, k) \equiv \angle (h, k) \quad \text{and} \quad \angle (h, k) \equiv \angle (k, h)$$

The first part of this axiom may be expressed in short as follows: Every angle can be uniquely placed in the given plane on a definite side of a given ray (Fig. 21).

(III.5) Let A, B, C be the three noncollinear points and A', B', C' be another three noncollinear points. If, in addition,

$$AB \equiv A'B', \quad AC \equiv A'C' \quad \text{and} \quad \angle BAC \equiv \angle B'A'C'$$

then

$$\angle ABC \equiv \angle A'B'C' \quad \text{and} \quad \angle ACB \equiv \angle A'C'B'$$

(Fig. 22).

A glance at the Group III axioms reveals that Axioms III.1-III.3 deal with segments only, Axiom III.4 deals with the congruence of angles, while Axiom III.5 deals with the congruence of segments in relation to that of angles.

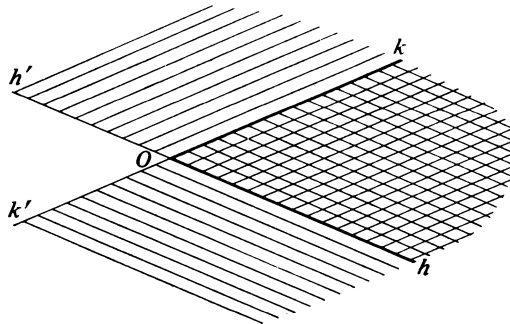


Figure 20

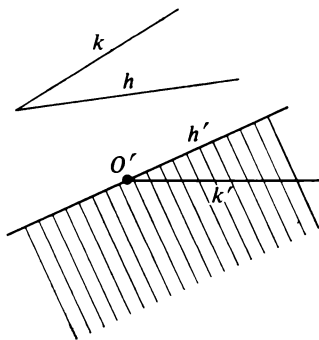


Figure 21

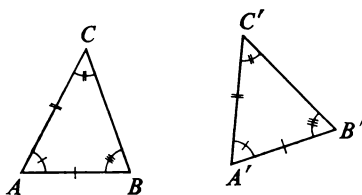


Figure 22

2.6. Corollaries of Axioms I-III

§. 17. We saw that the congruence of segments is mutual, that is, if some segment AB is congruent to segment $A'B'$, then $A'B'$ is congruent to AB . Therefore AB and $A'B'$ are called *mutually congruent*.

Take a system of points $A, B, C, \dots, K, L$ on a line a and another system of points $A', B', C', \dots, K', L'$ on line a' . If the segments AB and $A'B'$, AC and $A'C'$, BC and $B'C'$, $\dots$, KL and $K'L'$ are mutually congruent, the two systems are called *mutually congruent*. We have the following

Theorem 12. Assume that in two congruent systems $A, B, C, \dots, K, L$ and $A', B', C', \dots, K', L'$ the points of the first system are so situated that B lies between A on one side and $C, D, \dots, K, L$ on the other, C lies between A, B on one side and $D, \dots, K, L$ on the other, and so on. Then the points $A', B', C', \dots, K', L'$ are analogously situated, that is, B' lies between A' on one side and $C', D', \dots, K', L'$ on the other.

This can be formulated in short as follows: When a system of points is congruently moved from one line to another, the order of points is retained.

Omitting the proof of Theorem 12, we state the following theorem which will be frequently used later.

Theorem 13. Let A, B, C be three points on a line a and A', B', C' another three points on a line a' . Assume, also that $AB \equiv A'B'$ and $AC \equiv A'C'$. If B lies between A and C , and B' and C' on a' are on the same side of A' , then B' lies between A' and C' .

In contrast with Theorem 12, the congruence $BC \equiv B'C'$ is not assumed beforehand.

The theorem can be proved directly from Axioms III.1 and III.3. Indeed, by Axiom III.1, there is a point C^* on a' such that B' lies

between A' and C^* and $B'C^* \equiv BC$. Then $AC \equiv A'C^*$, by Axiom III.3. Thus $AC \equiv A'C'$ and $AC \equiv A'C^*$. But as C' and C^* lie on the same side of A' , Axiom III.1 implies that C' and C^* coincide. Hence B' lies between A' and C' . ■

Definition 8. A triangle ABC is said to be *congruent* to another triangle $A'B'C'$ if

$$\begin{aligned} AB &\equiv A'B', & AC &\equiv A'C', & BC &\equiv B'C' \\ \angle A &\equiv \angle A', & \angle B &\equiv \angle B', & \angle C &\equiv \angle C' \end{aligned}$$

(in notation $\triangle ABC \equiv \triangle A'B'C'$).

Theorem 14 (First theorem on the congruence of triangles). A triangle ABC is congruent to another triangle $A'B'C'$ if

$$AB \equiv A'B', \quad AC \equiv A'C' \quad \text{and} \quad \angle A \equiv \angle A'$$

Proof. By Axiom III.5, we have $\angle B \equiv \angle B'$ and $\angle C \equiv \angle C'$; thus we need only show that $BC \equiv B'C'$.

Assume that BC is not congruent to $B'C'$. Applying Axiom III.1, we can find a point D' on the ray $B'C'$ such that $BC \equiv B'D'$. Our assumption implies that the rays $A'C'$ and $A'D'$ are distinct. Applying Axiom III.5 to the triangles ABC and $A'B'D'$, we obtain $\angle BAC \equiv \angle B'A'D'$. But according to the given condition, $\angle BAC \equiv \angle B'A'C'$. The last two results contradict the uniqueness in Axiom III.4. Hence $BC \equiv B'C'$. ■

Theorem 15 (Second theorem on the congruence of triangles). A triangle ABC is congruent to a triangle $A'B'C'$ if

$$AB \equiv A'B', \quad \angle A \equiv \angle A' \quad \text{and} \quad \angle B \equiv \angle B'$$

The proof of this theorem is the same as that of the previous theorem (the "method of contradiction" is used together with Axioms III.1, III.5 and III.4).

What Axiom III.3 states for segments, the following theorem states for angles.

Theorem 16. Assume that h, k, l and h', k', l' are rays originating from O and O' , respectively, and that each of the triples of rays lies in one plane. Assume further that one of the rays h, k, l passes through the interior of the angle formed by the other two, and that the corresponding ray among the triple h', k', l' (the one denoted by the same letter) is situated analogously with respect to the other two. Then from

$$\angle(h, l) \equiv \angle(h', l') \quad \text{and} \quad \angle(l, k) \equiv \angle(l', k'),$$

it follows that

$$\angle(h, k) \equiv \angle(h', k')$$

Proof. We first prove this for the case where l passes through the interior of $\angle(h, k)$ (Fig. 23). Assume that $\angle(h, k)$ is not

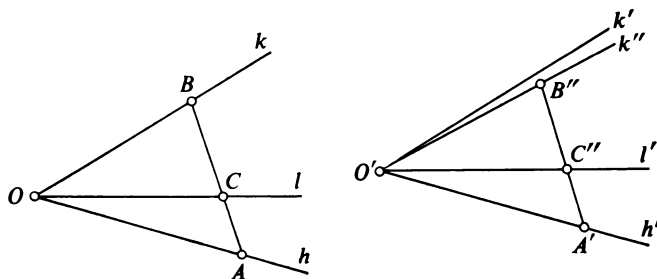


Figure 23

congruent to $\angle(h', k')$. On the basis of Axiom III.4, construct $\angle(h', k'')$ such that $\angle(h, k) \equiv \angle(h', k'')$ and that $\angle(h', k'')$ has interior points common to $\angle(h', k')$. Take points A and B on the rays h and k , respectively, and choose points A' and B'' on the rays h' and k'' so that $OA \equiv O'A'$ and $OB \equiv O'B''$. Then $AB \equiv A'B''$ by Theorem 14. Since l passes through the interior of $\angle(h, k)$, from Theorem 11a it follows that l intersects the segment AB at some point C . On the ray $A'B''$ take a point C'' such that $AC \equiv A'C''$. Due to the congruence relations $AC \equiv A'C''$ and $AB \equiv A'B''$ and Theorem 13, point C'' lies between A' and B'' ; moreover $BC \equiv B''C''$ (this can be proved by "contradiction", using Axiom III.3). Furthermore, noting that $OA \equiv O'A'$, $OB \equiv O'B''$, $\angle(h, k) \equiv \angle(h', k'')$ and using Axiom III.5, we have $\angle OAC \equiv \angle O'A'C''$ and $\angle OBC \equiv \angle O'B''C''$. The same axiom and the congruence relations $OA \equiv O'A'$, $AC \equiv A'C''$ and $\angle OAC \equiv \angle O'A'C''$ imply that $\angle AOC \equiv \angle A'O'C''$. Similarly, using $OB \equiv O'B''$, $BC \equiv B''C''$, $\angle OBC \equiv \angle O'B''C''$, we have $\angle COB \equiv \angle C''O'B''$.

Due to the first of the two conclusions and Axiom III.4, point C'' must lie on l' . In that case, $\angle COB \equiv \angle C''O'B''$ is equivalent to $\angle(l, k) \equiv \angle(l', k'')$. But by the hypothesis of the theorem, $\angle(l, k) \equiv \angle(l', k')$. Since, by our assumption, the rays k' and k'' are distinct, the last two results contradict Axiom III.4.

Now assume that k lies inside $\angle(h, l)$ and k' inside $\angle(h', l')$. Take points A and C respectively on h and l , and on h' and l' take points A' and C' so that $OA \equiv O'A'$, $OC \equiv O'C'$. Denote by B the point of intersection of the ray k and the segment AC and by B' the point of intersection of k' and $A'C'$ (the existence of these points is now ensured by the location of the rays). From the hypothesis of the theorem, Axiom III.5 and Theorem 15, it follows that $CB \equiv C'B'$, whence, noting that $CA \equiv C'A'$, we obtain $BA \equiv B'A'$. Thus, $OA \equiv O'A'$ and $BA \equiv B'A'$; we also see that

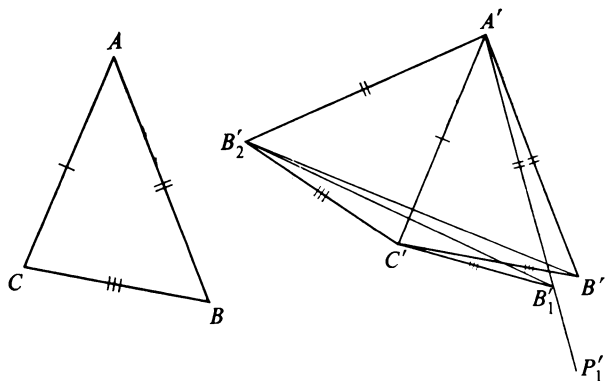


Figure 24

$\angle OAB \equiv \angle O'A'B'$ (according to Axiom III.5). Hence $\angle AOB \equiv \angle A'O'B'$. ■

The following theorem plays the same role for angles as Theorem 13 does for segments.

Theorem 16a. *Let the rays h, k, l and h', k', l' originating from O and O' , respectively, be given in a plane. Let k and l lie on the same side of the line containing h , while k' and l' lie analogously with respect to h' . If $\angle(h, k) \equiv \angle(h', k')$, $\angle(h, l) \equiv \angle(h', l')$ and if k lies in the interior of $\angle(h, l)$, then k' lies in the interior of $\angle(h', l')$.*

Proof. On the rays h and l take points A and C , respectively, and on h', l' take points A', C' such that $OA \equiv O'A'$ and $OC \equiv O'C'$. Since k passes through the interior of $\angle(h, l)$, it intersects segment AC at some point B . Using Theorem 14, Axiom III.1 and Theorem 13, we can easily show that there is a point B' on segment $A'C'$ such that $AB \equiv A'B'$. Further, applying Axiom III.5, we conclude that $\angle AOB \equiv \angle A'O'B'$. This together with Axiom III.4 implies that k' passes through B' . Thus k' lies in the interior of $\angle(h', l')$. ■

Theorem 17. *If a triangle ABC is such that $AC \equiv CB$, then $\angle CAB \equiv \angle CBA$ and $\angle CBA \equiv \angle CAB$.*

Proof. The assertion is obtained by applying Axiom III.5 to the triangles CAB and CBA . ■

Theorem 18 (Third theorem on the congruence of triangles). *A triangle ABC is congruent to triangle $A'B'C'$ if*

$$AB \equiv A'B', \quad AC \equiv A'C' \quad \text{and} \quad BC \equiv B'C'$$

Proof. Due to Theorem 14, we need only show that $\angle CAB \equiv \angle C'A'B'$. Let us assume the contrary. By Axiom III.4, there is a ray $A'P'_1$ which lies on the same side of $A'C'$ as point B' and which satisfies the condition $\angle CAB \equiv \angle C'A'P'_1$. By assumption, ray $A'P'_1$ does not coincide with ray $A'B'$ (Fig. 24).

By Axiom III.1, there is a point B'_1 on the ray $A'P'_1$ such that $AB \equiv A'B'_1$. Since $AB \equiv A'B'_1$, $AC \equiv A'C'$ and $\angle CAB \equiv \angle C'A'B'_1$, by Theorem 14 we have $\triangle ABC \equiv \triangle A'B'_1C'$. This implies the congruence $BC \equiv B'_1C'$. Since the congruence of segments is symmetric and transitive, we conclude from the previous result that the sides of $\triangle A'B'_1C'$ are congruent to the corresponding sides of $\triangle A'B'C'$. In the same fashion, we construct a triangle $A'B'_2C'$ lying on the opposite side of line $A'C'$ and having the same property. Consider the triangles $A'B'_2B'$ and $C'B'_2B'$. Since $A'B'_2 \equiv A'B'$, by Theorem 17 we have $\angle A'B'_2B' \equiv \angle A'B'B'_2$ and, similarly, $\angle B'B'_2C' \equiv \angle B'_2B'C'$. From these two results and Theorem 16, we conclude that $\angle A'B'_2C' \equiv \angle A'B'C'$, which together with Theorem 14 implies that $\triangle A'B'_2C' \equiv \triangle A'B'C'$. Hence, $\angle C'A'B'_2 \equiv \angle C'A'B'$. The proof that $\angle C'A'B'_2 \equiv \angle C'A'B'_1$ is exactly the same. The last two results contradict Axiom III.4 which proves the theorem. ■

Now it is easy to prove

Theorem 19. If $\angle(h, k) \equiv \angle(h', k')$ and $\angle(h, k) \equiv \angle(h'', k'')$, then $\angle(h', k') \equiv \angle(h'', k'')$.

Proof. Denote the vertices of $\angle(h, k)$, $\angle(h', k')$ and $\angle(h'', k'')$ by O , O' and O'' , respectively. On the rays h and k take points A and B (A on h and B on k) and on h' , k' , h'' , k'' choose points A' , B' , A'' , B'' such that $OA \equiv O'A'$, $OB \equiv O'B'$, $OA \equiv O''A''$ and $OB \equiv O''B''$. By Theorem 14, $AB \equiv A'B'$ and $AB \equiv A''B''$. Since the congruence of segments is symmetric and transitive, the previous relations yield $O'A' \equiv O''A''$, $O'B' \equiv O''B''$ and $A'B' \equiv A''B''$. This, in view of Theorem 18, implies that $\triangle O'A'B' \equiv \triangle O''A''B''$ and therefore $\angle A'O'B' \equiv \angle A''O''B''$. ■

Suppose now that a certain $\angle(h, k)$ is congruent to $\angle(h', k')$. Since by Axiom III.4 $\angle(h, k)$ is congruent to itself: $\angle(h, k) \equiv \angle(h, k)$, Theorem 19 shows that $\angle(h', k')$ is congruent to $\angle(h, k)$. Thus

$$\angle(h, k) \equiv \angle(h', k')$$

implies that

$$\angle(h', k') \equiv \angle(h, k)$$

This shows that the congruence relation for angles is (mutually) symmetric, and Theorem 19 shows that it is transitive as well. This in turn means that congruence relation for triangles is symmetric and transitive.

Further fundamental propositions of geometry can be obtained, for example, in the following order.

Definition 9. Two angles are called *adjacent* if they have a common vertex and one common side, while the remaining sides form a line. Two angles with a common vertex are called *vertical* if their sides form lines pairwise.

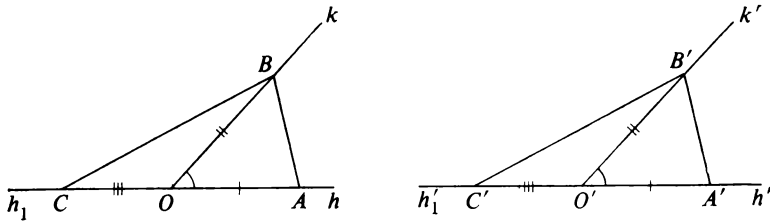


Figure 25

Theorem 20. *If two angles are mutually congruent, so are the adjacent angles.*

Proof. Suppose that $\angle(h, k) \equiv \angle(h', k')$ (Fig. 25). Denote by h_1 the ray which with h forms a line, by h'_1 the ray which with h' forms a line, by O and O' the vertices of $\angle(h, k)$ and $\angle(h', k')$, respectively. On h, k and h_1 take points A, B and C , respectively. By Axiom III.1, there are points A', B' and C' on h', k' and h'_1 such that $OA \equiv O'A', OB \equiv O'B'$ and $OC \equiv O'C'$. Therefore, $AC \equiv A'C'$ by Axiom III.3; $\angle OAB \equiv \angle O'A'B'$ (or $\angle CAB \equiv \angle C'A'B'$) by Axiom III.5; and $AB \equiv A'B'$ by Theorem 14. Since $AB \equiv A'B', AC \equiv A'C'$, and $\angle CAB \equiv \angle C'A'B'$, another application of Theorem 14 yields $BC \equiv B'C'$. Noting that $OB \equiv O'B', OC \equiv O'C'$ and $BC \equiv B'C'$, we see that Theorem 18 implies $\angle EOC \equiv \angle B'O'C'$, that is, $\angle(k, h_1) \equiv \angle(k', h'_1)$. ■

Theorem 21. *Vertical angles are congruent among themselves.*

This theorem follows easily from Theorem 20, since vertical angles have a common adjacent angle.

An angle which is congruent to its adjacent angle is called a *right angle*.

In order to prove the existence of right angles, take any $\angle(h, k)$ and construct $\angle(h', k)$ congruent to $\angle(h, k)$ but situated on the opposite side of k (Axiom III.4 ensures that such a construction is possible). On h and h' take congruent segments originating from the common vertex and join their endpoints by a line. If this line passes through the vertex of $\angle(h, k)$, then the angle $\angle(h, k)$ is a right angle; otherwise, this line intersects either the ray k or its complement. Then from Axiom III.5 or from Theorem 20 and Axiom III.5 it follows that this line forms right angles either with ray k or with its complement.

Theorem 22. *All right angles are congruent among themselves.*

Proof. Let $\angle(h, k)$ and $\angle(h', k')$ be right angles (Fig. 26), and let $\angle(k, h_1)$ and $\angle(k', h'_1)$ be their adjacent angles and O, O' their vertices. Assume that $\angle(h, k) \not\equiv \angle(h', k')$. According to Axiom III.4, there is a ray k'' originating from O' on the same side of the line (h'_1, h') as k' and is such that $\angle(h, k) \equiv \angle(h', k'')$. By our

assumption, k'' cannot coincide with k' . Then k'' must pass through the interior of either $\angle(h', k')$ or $\angle(h'_1, k')$ (this follows from Theorem 11a and Pasch's Axiom II.4). Suppose, for example, that k'' passes through the interior of $\angle(h', k')$. Then, by Axiom III.4, there is a ray k''_1 which originates at O' and lies on the same side of (h'_1, h') as k' and which is such that $\angle(h', k'') \equiv \angle(h'_1, k''_1)$. Since $\angle(h', k') \equiv \angle(h'_1, k')$, Theorem 16a implies that k''_1 lies inside $\angle(h'_1, k')$; accordingly, k''_1 does not coincide with k'' . This together with Axiom III.4 yields $\angle(h'_1, k'') \equiv \angle(h'_1, k''_1)$. But, on the other hand, $\angle(h'_1, k'') \equiv \angle(h'_1, k)$ (by Theorem 20) $\equiv \angle(h, k) \equiv \angle(h', k'') \equiv \angle(h'_1, k''_1)$, which contradicts the previous result. A similar contradiction is obtained if k'' passes through the interior of $\angle(h'_1, k')$ which proves the theorem. ■

Definition. Let A and B be two distinct points. The point O is called the *midpoint of the segment AB* if it lies on the line AB and is such that $AO \equiv OB$.

Theorem 23. *Every segment has a unique midpoint which is its interior point.*

In other words, every segment can be uniquely divided into two equal parts.

Proof. Consider the segment AB (Fig. 27). We construct congruent angles $\angle MAB$ and $\angle NBA$ such that the rays AM and BN lie on opposite sides of the line AB ; this is possible in view of Axiom III.4. On rays AM and BN take two congruent segments AC and BD . Since C and D lie on the opposite sides of line AB , segment CD intersects AB in some point; call this point O .

In choosing the segments AC and BD , we must take care of the following: if the lines AM and BN happen to intersect, we take C somewhere between A and the point of intersection of AM and BN , and then construct $BD \equiv AC$ (in reality this can never happen, but we shall not concern ourselves here with a proof of this). Now it is clear that O coincides neither with A nor with B . It is also easy to show that O cannot lie outside the segment AB . Namely,

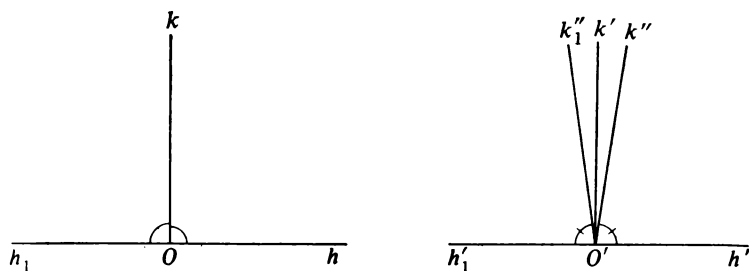


Figure 26

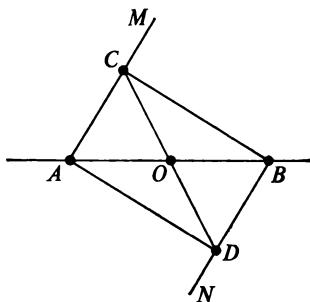


Figure 27

the supposition that A lies between O and B contradicts Pasch's axiom as applied to triangle OBD (line AM intersects the segment OB at A but intersects neither OD nor BD). Thus O lies between A and B . We shall show that O is the midpoint of segment AB . In fact, by Theorem 14, the triangles ABC and ABD are congruent, which yields $CB \equiv AD$. This result and Theorem 18 imply that the triangle ACD and BCD are congruent, which, in turn, means that $\angle ACD$ and $\angle CDB$ are congruent. This result with the aid of Theorem 15 shows that the triangles ACO and BDO are congruent, hence $AO \equiv OB$.

We shall now show that the midpoint of a segment is unique. Assuming otherwise, let the segment AB have two midpoints. In view of Axiom III.1, one of these lies between the other and the point A ;* call them O_1 and O_2 and let O_1 lie between A and O_2 . By Lemma 2, then, O_2 lies between O_1 and B . Noting that $AO_1 \equiv BO_1$ and $AO_2 \equiv BO_2$ as well as that O_1 lies between A and O_2 we see that Theorem 13 implies that O_1 lies between B and O_2 . Thus, on one hand we find that O_2 lies between B and O_1 , while on the other hand that O_1 lies between B and O_2 . This contradicts Axiom II.3. ■

We mention the following theorems.

Theorem 17a. *In an isosceles triangle, the median is also the altitude of the triangle and the bisector of the angle at the top vertex.*

Theorem 24. *Every angle can be uniquely divided into two equal parts.* ■

Theorem 25. *From a given point one and only one perpendicular can be drawn to a given line.*

Theorem 26. *At every point on a line one and only one perpendicular can be erected to this line.*

§ 18. The axioms of Groups I-III enable us to introduce the relations "greater than" and "less than" for segments and angles.

* By Axiom III.1, the midpoint of the segment lies inside it. This implies that if AB has two midpoints, one of them lies between the other and point A .

Definition 10. A segment AB is *greater than* segment $A'B'$ or $A'B'$ is *less than* AB if there is a point C inside AB such that

$$AC \equiv A'B'$$

(in notation, $AB > A'B'$ or $A'B' < AB$).

Definition 11. An angle $\angle(h, k)$ is *greater than* angle $\angle(h', k')$ or $\angle(h', k')$ is *less than* $\angle(h, k)$ if among the half lines passing through the vertex of $\angle(h, k)$ internally there is a half line l such that

$$\angle(h, l) \equiv \angle(h', k')$$

Theorem 27. For any two segments AB and CD one of the relations

$$AB \equiv CD, \quad AB > CD, \quad AB < CD$$

always holds, and one of them rules out the other two.

Indeed, by Axiom III.1, there is a point M on the line AB which lies on the same side of A as point B and $AM \equiv CD$. If M lies between A and B , then $AB > CD$; if M coincides with B , then $AB \equiv CD$; if B lies between A and M , then $AB < CD$. Thus we have shown that one of three relations always holds.

Now we shall show that one of them rules out the other two. Suppose, for example, that $AB > CD$. In this case, there is a point M on the segment AB such that $AM \equiv CD$. If AB and CD were to satisfy one more relation, say $AB \equiv CD$, in addition to $AB > CD$, then Axiom III.2 would imply that $AM \equiv AB$, which would contradict Axiom III.1. Similarly, $AB > CD$ and $AB < CD$ cannot hold simultaneously. Indeed, if they could, there would be a point M between A and B such that $AM \equiv CD$ and a point N between C and D such that $CN \equiv AB$. This contradicts Theorem 13. ■

Theorem 28. If $AB < A'B'$ and $A'B' < A''B''$, then $AB < A''B''$.

This result can be proved by obvious arguments using Theorem 13 and Theorem 8 (or Lemma 2). As a corollary of Theorem 28 we have

Theorem 29. If the segment CD is a part of the segment AB , then $CD < AB$.

The reader may easily formulate analogues of Theorems 27, 28 and 29 for the comparison of angles.

Having introduced the notions of "greater than" and "less than" for segments and angles, the following theorems may be formulated and proved.

Theorem 30. The exterior angle of a triangle is greater than each of its interior angles except the one adjacent to it.

Although Theorem 30 plays an important role in our exposition, we omit its proof here, since the proof which is usually given in textbooks can be justified by the axiom of Groups I-III. This theorem was mentioned in § 5, where its proof was also given.

Theorem 31. *In any triangle at least two of the angles are acute.*

Theorem 32. *In any triangle the larger side lies opposite the larger angle, and, conversely, the larger angle lies opposite the larger side.*

Theorem 33. *A perpendicular is shorter than an inclined line.*

Theorem 34. *Each side of a triangle is less than the sum and greater than the difference of the remaining two sides.*

Theorem 34 implies that a straight segment is shorter than a broken line joining the endpoints.

We have mentioned a number of theorems that can be proved on the basis of the axioms of Groups I-III. But at the same time many important geometrical facts cannot be deduced from these axioms. For example, these axioms do not imply that a line passing through an interior point of a circle intersects this circle. With the aid of the axioms of Groups I-III, it still cannot be shown, as was also the case using only the axioms of Groups I-II, that the set of geometrical elements is uncountable (for details in this connection, see § 72).

§ 19. The axioms of Group III enable us to define motion. As noted earlier, Euclid accepted motion as an obvious concept, not justified by any axiom, coinciding figures are considered equal. Thus in the Euclidean system the concept of motion is fundamental (although it remained unsubstantiated) and congruence is deduced from this. Hilbert introduced congruence as a basic concept with the aid of which the concept of motion can be defined. This definition will now be given.

Let Ω and Ω' be two sets of points, whether they are finite or infinite is immaterial. Let the points of these sets be in a one-to-one correspondence. Each pair of points M and N of Ω determines a segment MN . Let M' and N' be points of Ω' corresponding to the points M and N . We say that segment $M'N'$ corresponds to segment MN .

If the correspondence between Ω and Ω' is such that the corresponding segments are always mutually congruent, the sets Ω and Ω' are called *congruent*. We also say that each of the sets Ω and Ω' is obtained from the other by *motion* or that one of the sets can be *superposed* on the other. Corresponding points of the sets Ω and Ω' are said to *coincide through superposition*. (The difference between congruent proper and mutually reflecting sets is not introduced here.)

The following theorems hold.

Theorem 1. *Under motion, points of a line are mapped into points also lying on a line.*

This theorem is a direct consequence of Theorem 34. Indeed, consider a point set on a line a . We have to show that the points of the congruent set lie on one line a' . Choose three points A, B, C of the given set on a and, to be definite, assume that B is between

A and C . Thus the segment AC is made up of two segments AB and BC . If the points A', B', C' obtained from A, B, C by congruent movement do not lie on one line, then they form a triangle and, by Theorem 34, the segment $A'C'$ must be less than the segment composed of $A'B'$ and $B'C'$. Since $A'B' \equiv AB$ and $B'C' \equiv BC$, we must have $A'C' < AC$, which contradicts the fact that the two sets are congruent. ■

The following theorems are given without proof.

Theorem II. *Under motion, points of a plane are mapped into points of some plane.*

Theorem III. *The angle between two segments joining a point of the set with two other points of the same set is congruent to the angle between the corresponding segments of the congruent set.*

Theorem A. *Let M, N, P, Q be four noncoplanar points of some configuration Ω (that is, of some set of points), M' be any point in space, a be any line through M' , and α any plane through a . Then through motion Ω can be mapped into a position such that M coincides with M' , N lies on a preassigned side of M' , P lies in α on a preassigned side of a , and Q lies on a preassigned side of α .*

Theorem B. *If three noncollinear points M, N, P of a configuration Ω coincide with the corresponding points M', N', P' of a congruent configuration Ω' then two cases are possible: (i) every point of Ω coincides with the corresponding point of Ω' or (ii) every point of Ω lying in the plane MNP coincides with the corresponding point of Ω' while the remaining corresponding points of these configurations lie on either side of the plane MNP and every point of Ω' is uniquely determined by the position of the corresponding point of Ω (in the latter case the configurations are said to be symmetric or mutually reflecting in the plane MNP).*

In plane geometry Theorems A and B assume the following form.

Theorem C. *Let M, N, P be three noncollinear points of some configuration Ω , M' be any point of the plane, and a a line through M' . Then by motion Ω can be mapped into a position such that M coincides with M' , N lies on a preassigned side of M' and P lies on a preassigned side of a .*

Theorem D. *If two distinct points M and N of a configuration Ω coincide with the corresponding points M' and N' of the congruent configuration Ω' , then there are two possible cases: (i) every point of Ω coincides with the corresponding point of Ω' or (ii) every point of Ω lying on the line MN coincides with the corresponding point of Ω' while the remaining corresponding points of these configurations lie on either side of MN and each point of Ω' is uniquely determined by the position of the corresponding point of Ω (in this case the configurations are said to be symmetric with respect to MN).*

Theorems A and C characterize the degree of freedom of displacements of configurations. Theorems B and D establish conditions

determining the positions of configurations; namely, three points of a configuration determine its position in space to within reflection, while in a plane two points determine the position of configuration to within symmetry with respect to a line.

In defining the motion of a configuration Ω , we may assume, for one, that the set of its points fill the entire space and, in the case of plane geometry, the entire plane (that is, it may be assumed that every point in space and, in plane geometry, every point in the plane has a corresponding point, and if M', N' correspond to M, N , then $MN \equiv M'N'$). Then we say that the motion is performed of the entire space or, in the case of plane geometry, of the entire plane.

A motion of a configuration or of the entire space is called a *rotation about a point* O if O coincides with its corresponding point O' , that is, if O retains its position. A motion is a *rotation about a line* a if every point of a coincides with its corresponding point, that is, if every point of a retains its position. The line a is known as the *axis of rotation*.*

A motion is said to be a *translation along a line* u if the following three conditions hold:

- (1) every point of u is displaced but remains on u ;
- (2) every point of a plane α passing through u remains on α on the original side of u ;
- (3) every point not on α remains on the original side of α .

A motion in which all the points retain their positions is referred to as a translation along any line.

The rotation about a point and translation along a line are particular cases of motions. On the other hand, any motion can be achieved by a successive performance of translation and rotation.

In order to make the meaning of the last statement precise, we point out a fact which plays a fundamental role in the investigation of motions.

Suppose that two motions of space are performed one after another. The first of these shifts any point M to a point M' and the second shifts M' to M'' . This results in a transformation of all the points of space in which any point M is shifted to M'' . Such a transformation will be called the *product of motions*. In order to define the product of motions, it is not enough to mention its component motions; the order in which they are performed must also be indicated.

Theorem. *The product of two motions is a motion.*

The proof of this important theorem is obvious. Indeed, let any two points M and N in space be shifted to M' and N' by the first

* These definitions differ from the usually accepted ones, since we have not ruled out reflections of configurations.

of these motions and to M'' and N'' by the second motion. We must show that the segment MN is congruent to the segment $M''N''$. By the condition of the theorem, $MN \equiv M'N'$ and $M'N' \equiv M''N''$, whence Axiom III.2 and the propositions following from it imply that $MN \equiv M''N''$. ■

The property of motions expressed by the above theorem is known as the *group property*. Having noted the presence of group property, we may discuss the question of representing arbitrary motions as a product of some special simple motions. In particular (as noted above), every motion is a product of translation and rotation.

To prove this, consider any configuration Ω (whose points may fill the entire space) and assume that a motion transforms it to a congruent configuration Ω' . Let M be any point of Ω whose new position is M' .

Denote by Ω'' the configuration obtained from Ω through translation which shifts M to M' . Obviously Ω' and Ω'' are congruent, and a point of Ω'' corresponding to M' of Ω' coincides with M' . Therefore the motion through which Ω'' coincides with Ω' is a rotation about M' . Thus any shift of the configuration from Ω to Ω' is the product of translation from Ω to Ω'' and rotation as a result of which Ω'' coincides with Ω' .

2.7. Group IV: axioms of continuity

§ 20. By means of the axioms of Groups I-III we compared segments such that of any two segments one is either greater than, less than or equal to the other (Theorem 27 of § 18).

However, the axioms of Groups I-III are not sufficient to define the process of measuring, so that the relation of a given segment to a linear unit would be expressed by a definite number.

The validity of the process of measuring segments is expressed by the following Axiom IV.1, which is usually known as *Archimedes' axiom*. This axiom enables us (with the choice of a linear unit) to define uniquely for every segment a positive number called the *length of the segment*. To be able to show conversely the existence of a segment whose length is equal to any preassigned number, we require one more axiom.

Deviating from the exposition of Hilbert, we take such an axiom to be Axiom IV.2, which is nothing but the well-known Cantor's principle regarding nesting of intervals. In Hilbert's system, Axiom IV.2 is replaced by the axiom of completeness, which will be associated with Cantor's axiom in Chapter 4.

(IV.1—Archimedes' axiom) *Let AB and CD be any two segments. Then on the line AB there is a finite number of points $A_1, A_2, \dots, A_n$ which are so situated that A_1 lies between A and A_2 , A_2 lies between A_1 and*

A_3 and so on, and the segments $AA_1, A_1A_2, \dots, A_{n-1}A_n$ are congruent to CD and B lies between A and A_n (Fig. 28).

(IV.2—Cantor's axiom) *On any line a let there be given an infinite sequence of segments $A_1B_1, A_2B_2, \dots$ in which each of the successive segments lies inside the preceding one. Suppose further that, whatever the preassigned segment, there is a number n such that A_nB_n is less than this segment. Then there exists a point X of a which lies inside each of the segments A_1B_1, A_2B_2 , etc. (Fig. 29).*

This axiom immediately implies that *there is only one point X which lies inside each of the segments A_1B_1, A_2B_2 , etc.*

Indeed, if there were another point Y on a lying inside each of the segments $A_1B_1, A_2B_2, \dots$, then for any n the segment A_nB_n would be greater than XY , which contradicts the condition.

Definition 12. Suppose that to every segment there corresponds a definite positive number such that

- (1) to equal segments there correspond equal numbers;
- (2) if B is a point of the segment AC and numbers a and b correspond to the segments AB and BC , respectively, then $a + b$ corresponds to AC ;
- (3) to a segment OO' there corresponds the number 1.

The number which corresponds to a given segment in the above manner is called the *length* of this segment; the segment OO' is called the *linear unit* or the *unit length*.

We shall show that Conditions (1), (2), and (3) uniquely determine the length of every segment. First we shall assume that to each segment there corresponds a positive number such that Conditions (1)-(3) are fulfilled and shall show that there is no other correspondence between segments and numbers satisfying Conditions (1)-(3). Then we shall demonstrate that such a correspondence is possible. (In other words, first we shall show the uniqueness and then the existence of length.)

First of all note that if the segment AB is greater than the segment $A'B'$, then the length a of AB is greater than the length a' of $A'B'$. Indeed, according to the definition of "greater than" (see § 18), AB contains a point P which together with A defines a segment AP equal to $A'B'$. Denote by x and y ($x > 0, y > 0$) the lengths of AP and PB . By Condition (2), we have $a = x + y$, while by Condition (1) $x = a'$; whence $a = a' + y$ and, accordingly, $a > a'$.

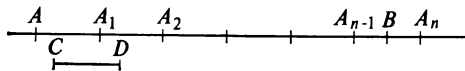


Figure 28

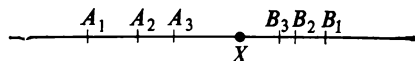


Figure 29

Further, according to Theorem 23, the linear unit OO' can be halved. Denote by O_1 the midpoint of segment OO' . Since the lengths of congruent segments OO_1 and O_1O' are equal and their sum is unity, each of these has the length $1/2$. Dividing OO_1 into two equal parts by a point O_2 , we find that the length of OO_2 is $1/2^2$, and so on. The segments OO_1 , OO_2 , ... will be called half, one-fourth, etc. of the linear unit.

Consider now an arbitrary segment AB with length a . From A towards B on the line AB mark off the segments AA_1 , A_1A_2 , etc., which are congruent to OO' . If one of the points A_n coincides with B , then by condition (2) $a = n$. If, however, no point A_1 , A_2 , ... coincides with B , then, according to Archimedes' axiom, there are two points A_n and A_{n+1} such that B lies between them. In this case a must satisfy the inequalities

$$n < a < n + 1$$

because AB is greater than AA_n and less than AA_{n+1} and the lengths of these two segments are, respectively, n and $n + 1$. Thus a is defined to within unity. We shall now show that a can be defined to within any accuracy. The process below which allows to find the value of a is called *measuring*.

Divide the segment A_nA_{n+1} by a point P_1 into two halves. Then B lies between either A_n and P_1 or P_1 and A_{n+1} or else coincides with P_1 . In other words, the segment A_nB is less than or greater than or equal to the linear unit. Therefore we have, respectively

$$n < a < n + 1/2$$

or

$$n + 1/2 < a < n + 1$$

or

$$a = n + 1/2$$

In the last case, a is determined exactly and the process of measuring is completed, while in the first two cases a is determined to within $1/2$ and the measuring is to be continued. Of the two segments A_nP_1 and P_1A_{n+1} dividing the one which contains B into two equal parts by a point P_2 , we may, depending on the position of B , either determine a exactly if B coincides with P_2 and stop the process or find a to within $1/4$ if B does not coincide with P_2 and continue the process in a similar fashion.

Instead of trying to enclose a between closer bounds, it is more convenient to obtain a binary representation for a :

$$a = n.n_1n_2 \dots$$

Here n is the integral part indicating the number of linear units contained in AB ; n_1 is the first place after the decimal point which is either 1 or 0 depending on whether or not AB contains, besides n linear units, half of the linear unit; n_2 is also either 1 or 0 depending on whether or not AB also contains, apart from n linear units and n_1 half of the linear unit, one-fourth of the linear unit and so on. The binary fraction expressing a may be either finite, if B coincides with one of $P_1, P_2, \dots, P_n, \dots$, which are constructed in the process of measuring, or infinite, if B does not coincide with any of these points. For example, if in the process of measuring we find that AB contains exactly one linear unit, one one-fourth and one one-eighth of the linear unit, then $a = 1.011$. In this case B coincides with P_3 . It is evident that a finite binary fraction can be formally looked upon as an infinite fraction; for example, $a = 1.011000\dots$. By saying that a binary fraction is infinite, we shall mean that it is essentially infinite, that is, it does not have a form in which after a certain digit only zeros follow. Assuming that to every segment there corresponds a certain length so that Conditions (1)-(3) are fulfilled, we have been able to find, with the aid of Archimedes' axiom, for any given segment every figure in the binary fraction representing this length. Thus the lengths of segments are uniquely determined by Conditions (1)-(3).

Now we must show that to every segment there corresponds a positive number so that Conditions (1), (2), (3) are fulfilled. To do this, with every segment as its length we associate the number obtained above in the process of measuring it. We have to show that then Conditions (1), (2), (3) are fulfilled.

First of all it is clear that the process of measuring as applied to the linear unit gives the number 1. Thus Condition (3) is fulfilled.

Further, for two congruent segments the measuring process gives equal values to their lengths. This is a direct consequence of Theorem 13, § 17, according to which the systems of points on two lines that are obtained in the measuring process have the same order of location; therefore, while measuring the two congruent segments in the resulting binary fractions there appear successively the same figures in analogous places. This shows that Condition (1) is also satisfied.

It remains to show that Condition (2) is satisfied.

First we shall prove two auxiliary propositions.

Auxiliary Proposition 1. *Given any segment PQ , it is always possible to choose a large number n such that by dividing the linear unit into 2^n equal parts we obtain segments each of which is less than PQ .**

* Every segment can be divided into 2^n equal parts, since every segment can be halved (see Theorem 23, § 17).

To prove this proposition, first assume that the linear unit OO' is divided by A into two equal parts, OA and AO' , and that each of them is greater than PQ . Then there is a point O_1 inside OA such that $OO_1 \equiv PQ$, and there is a point A_1 inside AO' such that $AA_1 \equiv PQ$. Mark off a segment $O_1O_2 \equiv PQ$ in the direction of A . Note that (i) A lies between O_1 and A_1 , and (ii) the congruence relations $O_1O_2 \equiv A_1A$ and $O_1A_1 \equiv A_1O_1$ hold; this together with Theorem 13 implies that O_2 lies between O_1 and A_1 . Applying Lemma 2 of § 14, we find that O_2 lies between O_1 and O' . Thus, if each half of the linear unit OO' is greater than PQ , by removing segments OO_1 and O_1O_2 , which are congruent to PQ , we do not cross the point O' . This implies that if for any n by dividing the linear unit into 2^n equal parts we obtain segments $\geq PQ$, then summing PQ as many times as we please, we cannot exceed the linear unit. This contradicts Archimedes' axiom, proving the proposition. ■

This proposition has a very important corollary. Namely, the process of measuring a segment cannot lead to an infinite binary fraction in which after a certain stage all the figures are 1's.

Indeed, suppose that a segment AB is being measured. Here we shall employ the same notation as the one used in describing the process of measuring above. If the measurement results in an infinite binary fraction with integral part n , then B lies between A_n and A_{n+1} . Assume first that in the fraction 1's appear successively immediately after the decimal point. Then B lies inside each of the segments P_1A_{n+1} , P_2A_{n+1} , ...; accordingly, for any n the segment BA_{n+1} is less than each of the 2^n equal parts of the linear unit, which contradicts Auxiliary Proposition 1. Assume now that the fraction has zeros in the k th place and then 1's repeat. Then B lies inside each of the segments $P_{k+1}P_k$, $P_{k+2}P_k$, ..., which is again in contradiction with Auxiliary Proposition 1.

The above fact simplifies the comparison of two binary fractions obtained by measuring segments. Namely, let a and b be two binary fractions obtained by measuring two segments. If up to certain places these fractions coincide and then in the next place a has zero and b unity, then one can assert that the number represented by a is less than the number represented by b (such an assertion may not be valid regarding arbitrary binary fractions; for example, the fractions 1.11000... and 1.10111... represent the same number).

Auxiliary Proposition 2. *If segment $A*B*$ is less than segment AB and numbers b^* and b are obtained by measuring these segments, then $b^* < b$.*

Since $A*B* < AB$, there is a point B' on AB such that $AB' \equiv A*B*$. We have to show that measurement of AB' yields a number less than the number obtained by measuring AB .

Starting from A we mark off towards B segments AA_1 , A_1A_2 , ... equal to the linear unit. Regarding any segment of line AB , let

us agree to say that a point belongs to the segment if it lies inside this segment or coincides with its left endpoint (assuming left to right is from A to B). For example, A_1 belongs to segment A_1A_2 , while A_2 to the next segment A_2A_3 . If, under this condition, B' and B belong to distinct segments of the system $AA_1, A_1A_2, \dots$, then the integral part of b^* is less than the integral part of b , and therefore $b^* < b$. But if both B' and B belong to the same segment A_iA_{i+1} , then b^* and b have the same integral parts. In this case, we divide the segment A_iA_{i+1} into two halves. If B and B' happen to be in different halves, then the first figure after the decimal in b^* is zero, while in b it is 1; accordingly, $b^* < b$. But if both B' and B lie in the same half of the segment A_iA_{i+1} , then, b^* and b have identical integral parts and the same figure after the decimal point. Then we divide that half of A_iA_{i+1} which contains both B' and B , and so on.

Continuing this process, we finally establish that $b^* < b$, provided B' and B never lie both in the same half obtained by dividing the segment. However, the last assumption must be rejected, because it means that the segment $B'B$ is less than each of the 2^n equal parts of the linear unit for any n , which contradicts Auxiliary Proposition 1. ■

We are now in a position to show that Condition (2) is also satisfied.

Let AC be any segment and B a point inside it. Denote by a, b, c the numbers obtained by measuring the segments AB, BC and AC . We must prove that

$$a + b = c$$

Fix a positive integer n and starting from B in the direction of A mark off segments $BA_1, A_1A_2, \dots$ congruent to the segments obtained by dividing the linear unit into 2^n equal parts. Archimedes' axiom implies that among the points $A_1, A_2, \dots$ there are two points, A_k and A_{k+1} , such that A_k either belongs to BA or coincides with A and A_{k+1} in combination with B determines a segment BA_{k+1} containing A . Similarly, determine points C_l and C_{l+1} marking off segments $BC_1, C_1C_2, \dots$ in the direction of C that are congruent to the segments A_iA_{i+1} . Evidently, the following inequalities hold between the segments:

$$BA_k \leq AB_1' < BA_{k+1}, \quad BC_l \leq BC < BC_{l+1} \\ A_k C_l \leq AC < A_{k+1} C_{l+1}$$

From this and noting Auxiliary Proposition 2, these inequalities yield inequalities for the corresponding numbers:

$$\frac{k}{2^n} \leq a < \frac{k+1}{2^n}, \quad \frac{l}{2^n} \leq b < \frac{l+1}{2^n}, \quad \frac{k+l}{2^n} \leq c < \frac{k+l+2}{2^n}$$

These inequalities yield

$$\frac{k+l}{2^n} \leq a + b \leq \frac{k+l+2}{2^n}, \quad \frac{k+l}{2^n} \leq c < \frac{k+l+2}{2^n}$$

hence

$$|a + b - c| < \frac{1}{2^{n-1}}$$

Since n is any positive integer, it follows that

$$a + b - c = 0$$

or $a + b = c$, as required.

Thus the axioms of Groups I-III and Axiom IV.1 enable us to define the measurement of segments and with each segment to associate a positive number called its length. Length is uniquely defined by Conditions (1)-(3).

According to Condition (1), equal segments have equal lengths. Theorem 27 of § 13 and Auxiliary Proposition 2 imply also the converse, that is, segments having the same lengths are equal. In this manner, the comparison of segments can be replaced by that of corresponding lengths.

Similar to the definition of the length of a segment is the definition of the magnitude of an angle.

Definition 13. To every angle let there correspond a positive number such that the following conditions hold:

- (1) Equal numbers correspond to equal angles.
- (2) If a ray l passes inside $\angle (h, k)$ through its vertex and numbers α and β correspond to $\angle (h, l)$ and $\angle (l, k)$, then $\alpha + \beta$ corresponds to $\angle (h, k)$.
- (3) Number 1 corresponds to a certain angle $\angle (o, o')$.

Then the number that corresponds to a given angle in the above manner is called its *magnitude* and $\angle (o, o')$ is called the *angular unit*.

The uniqueness and existence of the magnitude of angles is proved in the same way as done for the length of segments. In doing so there is no need to introduce a new axiom analogous to Archimedes' axiom for segments; the corresponding proposition can be proved.

§ 21. In accordance with the previous result, apart from the set of all segments, the set of their lengths is a completely defined set; of course, we assume that a linear unit has been chosen. But from the Axioms of Groups I-III and Axiom IV.1 it does not follow that the lengths of segments exhaust all the real numbers. On the basis of these axioms it cannot even be established whether the set of lengths is uncountable.

Only if we extend the system of axioms by adjoining, for example, Cantor's axiom IV.2 can we prove the following

Theorem 35. *For any real number $a > 0$, there is a segment with length a .*

To prove this theorem, represent a in the form of a binary fraction $n.n_1n_2\dots$. Assume first that a cannot be represented by a finite binary fraction. Then the fraction $n.n_1n_2\dots$ cannot have, after a certain stage, only 1's (because the infinite fraction $n.n_1n_2\dots n_k0111\dots$ represents the same number as the finite fraction $n.n_1n_2\dots n_k1$).

Take a half line with vertex at the point A and on this half line construct segments $AA_1, A_1A_2, \dots, A_nA_{n+1}$ congruent to the linear unit. Divide the last of these segments, segment A_nA_{n+1} , into two equal parts by a point P_1 . We call that half of A_nA_{n+1} the left half which is situated on the same side as A , and the other half the right half. This terminology applies to any other segment of the half line if that segment is to be divided into two halves. Denote by l_1 the segment coinciding with the left half of A_nA_{n+1} if $n_1 = 0$ and with the right if $n_1 = 1$. We next divide l_1 into two halves by P_2 and denote by l_2 its left or right half depending on whether $n_2 = 0$ or $n_2 = 1$. This process is continued indefinitely. Thus we construct a sequence of segments $l_1, l_2, \dots$.

By construction, the interior points of each of these segments lie inside the preceding segment and one endpoint of a segment coincides with the endpoint of the preceding segment. It cannot happen, however, that, starting from a certain number all the segments l_n have a common endpoint (since the fraction $n.n_1n_2\dots$ cannot have only zeros or 1's after a certain stage). Therefore among the segments $l_1, l_2, \dots$ there is a segment l_{k_1} lying strictly inside l_1 , there is a segment l_{k_2} lying strictly inside l_{k_1} , and so on. Furthermore, Auxiliary Proposition 1, which was used to prove the existence of length, implies that there is no segment less than all of the segments $l_1, l_{k_1}, l_{k_2}, \dots$. Therefore we may apply Cantor's axiom IV.2 to the sequence $l_1, l_{k_1}, l_{k_2}, \dots$ and, according to this axiom, assert that there is a unique point B lying inside all the segments $l_1, l_{k_1}, l_{k_2}, \dots$. Clearly, this point B also lies inside all the segments $l_1, l_2, l_3, \dots$. Evidently, segment AB has the length a . In fact, measurement of AB gives just a .

Thus the theorem is proved when a cannot be represented by a finite binary fraction. If a is represented by a finite binary fraction, then the endpoint B of the required segment will be one of the points $A_n, P_1, P_2, \dots$. We omit the detailed arguments in this case and note only that here Cantor's axiom need not be used. ■

A proposition analogous to Theorem 35 holds for the magnitude of angles also. Namely, we have

Theorem 36. *Suppose that, with a choice of unit of measurement, the right angle has a magnitude ω . Then to any number α , $0 < \alpha < 2\omega$, there corresponds an angle whose magnitude is α .*

It is customary to choose the unit of measurement of angles so that the magnitude of the right angle is the number $\pi/2$. In this case the unit of measurement is called the *radian*.

Having defined the measurement of segments and angles and established, in Theorems 35 and 36, that a segment can be constructed with a given length and an angle with a given magnitude, we can apply arithmetic and algebra to geometry.

For example, employing the methods of arithmetic, we can prove the following important theorem.

Theorem 37. *There are points inside every segment that divide it into n equal parts.*

Indeed, let a segment AB be given. We have proved that every segment has some length; denote by a the length of AB . Using division of numbers, we obtain the number a/n . From Theorem 35 and Axiom III.1 it follows that on AB there exist segments $AA_1, A_1A_2, \dots, A_{n-2}A_{n-1}$ with the same length a/n . Obviously, the points $A_1, A_2, \dots, A_{n-1}$ are the points sought.

A similar theorem holds for angles.

Theorem 38. *Inside every angle through its vertex there pass half lines dividing this angle into n equal parts.*

§ 22. With the aid of the axioms of Groups I-IV, we may introduce the principle of coordinates for a line, a plane and space.

We first construct a coordinate system for a line. Take any line a , and on it choose a point O called the *origin*. Let us agree to call one of the two half lines defined on a through O *positive* and the other *negative*. Furthermore, we choose a segment as the *unit of measurement*.

With every point M on a we associate its *coordinate* x , taking the absolute value of x to be the length of OM and determining the sign from the position of M in the following manner: $x > 0$ if M lies on the positive half line, $x < 0$ if M lies on the negative half line, and $x = 0$ if M coincides with O . As an immediate consequence of Theorem 35, we have the following proposition.

For any number x there is just one point on the line with coordinate x .

We now introduce a system of coordinates in a plane. Let α be any plane, O any point in this plane, and a be some line in α through O . Then O divides a into two half lines, one of which is called *positive* and the other *negative*. Line a divides α into two half planes, one of which is called *positive* and the other *negative*. Choosing a unit of length, we define on a , in accordance with the previous process, a system of coordinates with the origin at O and a positive half line.

Take any point M in α . By Theorem 25 of § 17, a unique perpendicular can be drawn to a from M . Let the foot of this perpendicular be M_x . Let x be the coordinate of M_x in the system of coordinates

introduced on line a , and let y be a number whose absolute value is equal to the length of the segment MM_x . The sign of y is determined as follows: $y > 0$ if M lies in the positive half plane, $y < 0$ if M lies in the negative half plane, and $y = 0$ if M lies on a .

Thus to every point M in α there corresponds a pair of numbers x and y , called the *coordinates* of this point.

Obviously, *for every pair of real numbers x and y there is just one point in the plane with these numbers as its coordinates.*

Indeed, the number x always uniquely defines M_x on a . By Theorem 26 of § 17, at point M_x we can erect a unique perpendicular to a . Assume $y \neq 0$. Then, by Theorem 35, there is a segment whose length equals the absolute value of y . Starting from M_x mark off this segment on the perpendicular to a in such a way that it lies in the positive half plane if $y > 0$ and in the negative half plane if $y < 0$. Denote by M the endpoint of the segment thus marked off. Then M will have the given coordinates x and y .

When $y = 0$, assume that M coincides with M_x . Then M has the given coordinates x and $y = 0$.

Thus it is always possible to find a point whose coordinates are the given numbers x and y . The uniqueness of this point can be shown by obvious arguments.

To introduce coordinates in space, choose any plane α and define in it by some method a coordinate system (that is, take a point O , a line a , etc.). The plane α divides the space into two half spaces, one of which we shall call *positive* and the other *negative*. Then with every point M in space we associate three coordinates (x, y, z) defining them as follows: x and y coincide with the coordinates of M' (the foot of the perpendicular drawn from M to α) in the coordinate system that we assume to be introduced in α ; z equals, in absolute value, the length of the segment MM' ; $z > 0$ if M lies in the positive half space, $z < 0$ if M lies in the negative half space, and $z = 0$ if M lies in α .

This method of introducing coordinates in space requires prior definition of the perpendicular to a plane and the proof of the theorem which states that from any point there can be drawn only one perpendicular to the given plane. The required definition as well as the proof of the theorem can be given in the same way as is done in textbooks on elementary geometry.

If we also prove the existence and uniqueness of a perpendicular to a plane at a given point of it, then, using Theorem 26 of § 17, we can prove the following assertion.

For any three real numbers x, y, z there is just one point in space with the coordinates x, y, z .

The above systems of coordinates in plane and space could be termed Cartesian. It should be pointed out, however, that many properties which are characteristic of the Cartesian system do not

follow from the axioms of Groups I-IV alone. Consider for simplicity a coordinate system in the plane. Let us call, as is usually done, a as the x -axis and the line perpendicular to a at O as the y -axis. Take any point M in the plane and denote by M_x and M_y the feet of the perpendiculars drawn from M to x - and y -axis, respectively. Using the axioms of Groups I-IV we cannot show, for example, that the segment OM_y is equal to the segment M_xM . Similarly, from these axioms one cannot deduce the well known result of analytic geometry concerning the distance between two points.

§ 23. By introducing coordinates on a line, we establish a one-to-one correspondence between the set of all points of this line and the set of all real numbers. Let us devote some space to the special features of this correspondence.

As shown earlier by means of the axioms of Groups I-II, the set of points on the line can be ordered in such a way that if B follows A and precedes C , then B lies between A and C in the sense of § 13. Such an order can be established in two different ways only; this corresponds to our visual evidence of two directions on a line. Among these two orderings we choose the one in which origin precedes all the points of the positive half line (the direction defined by this will be called positive). Then, if M_1 and M_2 are two points with coordinates x_1 and x_2 and M_1 precedes M_2 , then $x_1 < x_2$.

Thus we have

Theorem 39. *Between the ordered set of all points of a line and the ordered set of all real numbers a one-to-one correspondence can be established such that the corresponding elements are situated in the same order.*

The property of the line expressed by this theorem is known as the *continuity property*. Since Axioms IV.1-IV.2 ensure the continuity of the line, they are referred to as *axioms of continuity*.

Axioms IV.1 and IV.2 can be replaced by equivalent statements, the axioms of Groups I-III remaining unchanged. One of the important statements equivalent to the axioms of Group IV is *Dedekind's principle*.

Well known in analysis is the proposition expressing Dedekind's principle for the set of real numbers.

If all the real numbers are divided into two classes such that (a) each number belongs to one and only one class and each class contains numbers and (b) each number of the first class is less than any number of the second, then either the first class contains the largest number or the second contains the smallest number.

This proposition rules out the following two possibilities: the presence of closing elements in both classes and the absence of closing elements in both classes.

Theorem 39 and Dedekind's principle for real numbers immediately yield Dedekind's principle for a line.

Theorem 40. *If all the points on a line are divided into two classes such that (a) each point belongs to one and only one class and each class contains points and (b) each point of the first class precedes every point of the second, then either in the first class there is a point which is preceded by all the other points of the first class or in the second class there is a point preceding all the other points of the second class.*

This point is said to define Dedekind's cut of the line.

That this statement is equivalent to the axioms of Group IV is proved by

Theorem 41. *If Dedekind's principle is adjoined to the axioms of Groups I-III, then Archimedes' axiom IV.1 and Cantor's axiom IV.2 can be proved.*

First we shall obtain Archimedes' axiom from Dedekind's principle and the axioms of Groups I-III.

We shall use the method of contradiction. Assume that the assertion of Archimedes' axiom does not hold for a certain segment AB . This means that there is an infinite sequence of congruent segments $AA_1 \equiv A_1A_2 \equiv \dots \equiv A_nA_{n+1} \equiv \dots$ lying inside AB .

We choose an order of points on the line AB (direction) in which A precedes B and divide the points of AB into two classes as follows: to the first class assign every point which precedes some point A (which means that it also precedes A_{n+1} , A_{n+2} , etc.) and to the second assign all the other points of AB .

Obviously, this division satisfies the conditions of Dedekind's cut. Indeed, (i) each point of AB belongs to one and only one class and each class is nonempty, since the first contains A_1 , A_2 , ..., A_n , ... and the second contains point B , and (ii) all the points of the first class precede the points of the second class.

Then Dedekind's principle, taken to be an axiom now, implies that there is a point C performing this cut. Since the first class evidently does not have any last element, C belongs to the second class and precedes all the other points of this class. By Axiom III.1, there is a point D which precedes C and determines together with it a segment CD congruent to each of the segments AA_1 , A_1A_2 , etc. The point D cannot belong to the second class because it precedes C . Thus D belongs to the first class and hence precedes some point A_n . The segment A_nA_{n+1} is a part of CD and, by Theorem 29, $A_nA_{n+1} < CD$. On the other hand, we have $A_nA_{n+1} \equiv CD$. But due to Theorem 27, both $A_nA_{n+1} < CD$ and $A_nA_{n+1} \equiv CD$ cannot hold simultaneously. This contradiction proves Archimedes' principle (or axiom).

We shall now obtain Cantor's axiom.

Let an infinite sequence of segments A_1B_1 , A_2B_2 , ... be given on a line a where each segment $A_{n+1}B_{n+1}$ lies inside A_nB_n . Assume

that there is no segment less than all the segments of the sequence. We have to show that there is a point which lies inside each segment $A_n B_n$.

Define a direction on the line and assume that A_n always denotes the endpoint of a segment preceding the endpoint B_n . Divide the points on a into two classes, assigning a point to the first class if it precedes one of the points A_n (consequently, it also precedes A_{n+1} , A_{n+2} , etc.) and assign the remaining points of a to the second class.

We obtain a Dedekind's cut. Indeed, (i) each point of a belongs to one and only one class; each class is nonempty, since the first contains the points $A_1, A_2, \dots, A_n, \dots$ and the second the points $B_1, B_2, \dots, B_n, \dots$, and (ii) the points of the first class precede the points of the second. By Dedekind's principle, there is a point C performing this cut.

Since the first class evidently does not contain a last element, C is the first point of the second class. Therefore C precedes all the points $B_1, B_2, \dots, B_n, \dots$ and follows each of the points $A_1, A_2, \dots, A_n, \dots$. From here we conclude that C lies inside each segment $A_n B_n$. ■

§ 24. We have seen that the axioms of continuity enable us to prove that a coordinate system may be introduced on every line and hence turn the line into a number axis.

This fact is very important, as it makes it possible to apply the basic results of analysis to geometry.

We give two theorems which can be easily proved, once the axioms of continuity have been introduced.

Theorem 42. *If a line passes through an interior point of a circle, this line intersects the circle in two points.**

Theorem 43. *If a circle k passes through an interior and an exterior point of another circle k' , the circles k and k' intersect at two points.*

We shall prove only the first theorem.

Consider a line a passing through an interior point of the circle k with radius r . From the centre of k draw a perpendicular to a and denote the foot of the perpendicular by O .

Introduce a coordinate system on a with O as the origin. The distance between any point on a with the coordinate x and the centre of the circle is a function of x , which we denote $s(x)$. It is easily seen that $s(x)$ is continuous for all x . Indeed, by a well-known theorem which states that the difference of two sides of a triangle is less than the third side we have

$$|\Delta s| = |s(x + \Delta x) - s(x)| < |\Delta x|$$

whence $\lim \Delta s = 0$ as $\Delta x \rightarrow 0$. Noting further that $s(0) < r$ and $s(r) > r$, we find that the function $s(x) - r$ changes its sign

* A point is an *interior* point of a circle if its distance from the centre is less than the radius and it is *exterior* if the distance is greater than the radius.

as x changes from 0 to r . Since this function is continuous, there is a value of x , say $x = x_1$, lying between 0 and r for which $s(x_1) = r$. Using the property of the continuity of a line, we may assert that for any number x_1 there is a point M_1 on a whose coordinate is x_1 (this was shown in § 23). Since the distance of M_1 from the centre of the circle is r , this point lies on the circumference of the circle, that is, it is a point of intersection of a and circle k .

It is easy to see that M_2 with the coordinate $x_2 = -x_1$ is the second point of intersection. ■

Theorems 42 and 43 justify the constructions usually employed in textbooks on elementary geometry when solving the problem of dividing a segment or an angle into two halves, of drawing from a given point a perpendicular to a given line, etc. In Theorem 23 on the question of dividing a segment into two halves we avoided using these constructions because without using the axioms of continuity Theorems 42 and 43 cannot be derived from the axioms of Groups I-III.

2.8. Group V: the axiom of parallelism.

Absolute geometry

§ 25. *Definition 14.* Two lines lying in one plane and having no common point are called *parallel*.

This definition evidently requires the proof of the existence of parallel lines. Following Euclid, this can be done by the help of the following theorem.

Theorem 44. *Let the lines a , b , c lie in one plane. If c intersects a and b such that the interior alternate angles formed by c with a and b are equal, then a and b are parallel.*

Theorem 44 can be proved at once by the method of contradiction. Let c intersect a and b at points A and B , respectively, and assume that a and b are not parallel. Then they have a common point O , and the exterior angle in the triangle AOB is equal to one of the remote interior angles. This contradicts Theorem 30. ■

A particular case of Theorem 44 is

Theorem 45. *Two lines lying in one plane and perpendicular to a third line are parallel to each other.*

Theorems 44 and 45 immediately imply

Theorem 46. *Through any point not on a given line there passes a line which is parallel to the given line.*

Indeed, let A be any point not on the given line a . Draw from A the perpendicular AP to a and let b be the line which passes through A . Then b is perpendicular to AP and lies in the plane containing AP and a . By Theorem 45, b is parallel to a .

Theorem 46 supplements Definition 14, as it establishes the existence of parallel lines.

In order to justify the Euclidean theory of parallels, it is enough to adjoin the following Axiom V to the axioms of Groups I-IV.

(V—the parallelism axiom) *Let a be any line and A a point not on a . Then in the plane determined by the point A and the line a one can draw at most one line through A that does not intersect a .*

In § 5 it was shown that this axiom is equivalent to Euclid's fifth postulate.

Axiom V immediately implies a converse to Theorem 44.

Theorem 47. *If two parallel lines are intersected by a third line, the alternate interior angles are equal.*

From this, in the usual manner one can deduce

Theorem 48. *The sum of the interior angles of a triangle is equal to two right angles.*

No purpose will be served by mentioning further theorems of geometry. All the arguments employed in textbooks in proving these theorems are justified by what has been described above and can be carried out without any appeal to diagrams or visual evidences.*

Note also that the axioms of Groups I-V substantiate Cartesian analytic geometry. In § 22 coordinate systems were introduced on a line, a plane and in space. Now, possessing Axiom V and hence the Euclidean theory of parallels, the theory of similarity of figures and, in particular, the theorem of Pythagoras, we can show that the distance between two points $M_1(x_1, y_1, z_1)$ and $M_2(x_2, y_2, z_2)$ is given by the formula

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

that a plane is defined by the equation of first degree

$$ux + vy + wz + t = 0$$

and so on. This makes it possible to prove geometrical theorems by means of arithmetics.

§ 26. In Chapter 1 we gave an account of various attempts to prove Euclid's parallel postulate. The authors of these proofs tried to deduce in a logical way the fifth postulate from the remaining postulates of Euclid. It should be noted that though this problem stood before geometers for many centuries, it remained undefined until the end of the nineteenth century.

Indeed, the definitions and axioms of Euclid were so far from perfection that they could hardly serve as a basis for developing strict-

* In fact, we assert here that the system of Hilbert's axioms is complete, that is, assuming all his axioms, it is possible to develop geometry in a strictly logical order. The precise definition of the completeness of a system of axioms and its proof are given in Chapter 4.

ly logical constructions. Interestingly, the fifth postulate problem, although solved by Lobachevsky, was yet to be formulated precisely, since in his time the flaws in Euclid's development of geometry had not been removed.

Having given an exposition of Hilbert's axioms, we may precisely formulate the fifth postulate problem in the following manner.

Assuming the axioms of Groups I-IV, derive Axiom V from them.

Lobachevsky's result can also be now expressed precisely.

Axiom V is not a consequence of the axioms of Groups I-IV.

This result may also be formulated as follows.

If to the axioms of Groups I-IV is adjoined a statement negating the truth of Axiom V, then the result of all these statements will be a logically consistent system (non-Euclidean geometry).

Basic facts of the theory of parallels of Lobachevskian geometry and the proof of its consistency are dealt with in Chapter 3.

§ 27. The system of propositions resulting from only the axioms of Groups I-IV is known as *absolute geometry* (J. Bolyai's terminology). Absolute geometry obviously forms the common portion of Euclidean and non-Euclidean geometries, since propositions proved with the aid of the axioms of Groups I-IV are equally valid in both Euclidean and Lobachevskian geometries.

All the theorems up to Theorem 46 formulated in this chapter are theorems of absolute geometry. To this list we add the theorems obtained by Saccheri, Lambert and Legendre that were proved in § 8.

Theorem 49. *The defect $D(\triangle)$ of any triangle satisfies the inequality*

$$D(\triangle) \geq 0$$

In other words, the sum of the angles of a triangle does not exceed two right angles.

Theorem 50. *The summit angles of Saccheri's quadrilateral cannot be obtuse (that is, the obtuse angle hypothesis leads to a contradiction).*

Theorem 51. *If there is one triangle with positive defect, every triangle has positive defect. In other words, if there is one triangle the sum of whose angles is less than two right angles, the sum of the angles of every triangle is less than two right angles.*

Theorem 52. *If the acute angle hypothesis is assumed true for one Saccheri's quadrilateral, it is true for every Saccheri's quadrilateral.*

Theorem 53. *Saccheri's right angle hypothesis and Legendre's assumption that there exists a triangle the sum of whose angles is two right angles are equivalent to Axiom V.*

Theorem 54. *If there is an acute angle such that the perpendicular erected at any point of one side meets the other side, then the assertion of Axiom V can be proved.*

Chapter 3

The Non-Euclidean Theory of Parallels

3.1. Lobachevsky's definition of parallels

§ 28. We shall now deal with the basic facts of the non-Euclidean theory of parallel lines. This theory is based upon the axioms of Groups I-IV of absolute geometry and the following the Lobachevskian parallel axiom.

There exist a line a and a point A , not on a , such that through A there pass at least two lines not intersecting a and lying in one plane with a .

Let us show that in the same plane there pass infinitely many lines through A not intersecting a .

Denote by a_1 and a_2 two lines through A that do not intersect a (Fig. 30); the existence of these lines is guaranteed by the Lobachevskian parallel axiom. Choose a point B_2 on a_2 such that this point and a are on the opposite side of a_1 . Join B_2 with some point B of a . The segment B_2B intersects a_1 at some point of B_1 . Denoting any point of B_1B_2 by M , we easily find that the line AM does not intersect a . In fact, if the line AM intersects a at a point C lying in the direction of M from A , a triangle MBC is formed whose side MB is intersected by a_1 . Then, by Pasch's axiom II.4, line a_1 must intersect a , which is impossible. If, however, line AM intersects a at a point C' lying in the direction of A from M , a triangle MBC' is formed whose side MC' is intersected by a_2 . Then Pasch's axiom II.4 implies that a_2 must intersect a , which is also impossible.

Thus we can conclude that if a_1 and a_2 pass through A and do not intersect a , then all the lines through A constituting a definite pair of vertical angles formed by a_1 and a_2 do not intersect a .

The Lobachevskian axiom negates the characteristic property of Euclidean geometry concerning the uniqueness of a parallel line at

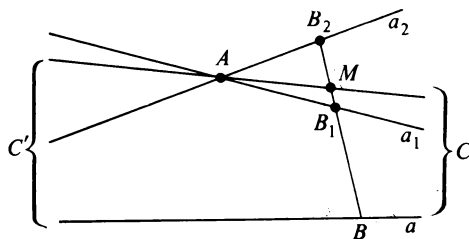


Figure 30

least for a certain point and a certain line. However, it is easily seen that

if the relationship asserted by the Lobachevskian axiom is true for a certain point and a certain line, it is also true for any point and any line.

We shall prove this by the method of contradiction.

Suppose that, contrary to our assertion, through some point B not on the line b there passes only one line b' not intersecting b and lying with it in one plane (Fig. 31). Draw a perpendicular BB_1 from B to b and take another point B_2 , distinct from B_1 , on b .

It is easy to show that the alternate interior angles formed by BB_2 with b and b' are equal. Indeed, if it were not so it would be possible to draw through B a line b'' distinct from b' such that BB_2 forms with b and b'' equal alternate interior angles. But then, on the one hand, b'' would not have a common point with b , contradicting Theorem 44 of Chapter 2, and on the other, it would meet b , since the only line passing through B and not intersecting b is, by our assumption, line b' . Similarly, the line BB_1 , too, forms with b , b' equal alternate interior angles; accordingly, BB_1 is perpendicular not only to b but also to b' . These results immediately imply that for $\triangle BB_1B_2$ the sum of the interior angles is two right angles. Then Theorem 51 of Chapter 2 implies that so is the sum of the angles of any triangle. Taking into account Theorem 53, we see that this in turn implies the fifth postulate of Euclid and hence that the line passing through a given point and not intersecting a given line is unique. This contradicts the Lobachevskian axiom negating this uniqueness with regard to line a and point A .

Thus assuming the validity of Lobachevsky's axiom we must assert the following proposition.

Theorem I. *Given any line and any point not on this line, through the given point there pass infinitely many lines not intersecting the given line.*

Here we are concerned of course with lines lying in one plane with the given line. This condition will not be specified further and our

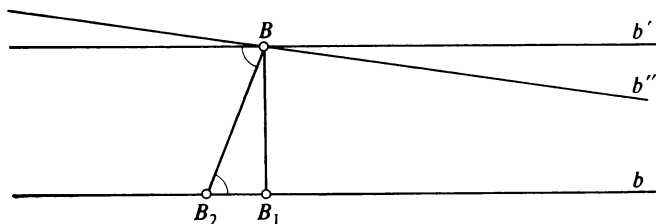


Figure 31

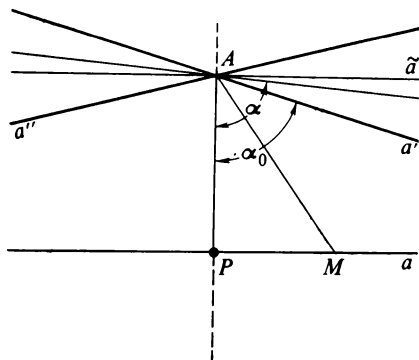


Figure 32

discussion (up to § 33) will concern plane geometry, that is, points and lines of a definite plane.

According to the aforementioned fact, we must accept, as a corollary to the Lobachevskian axiom (i) the acute angle hypothesis for every Saccheri's quadrilateral and (ii) that the sum of the angles of every triangle is less than two right angles.

§ 29. In contrast with Euclid's definition of parallel lines, according to Lobachevsky's definition only some special lines, among many not intersecting the given line, are taken to be parallel to the given line. We shall now present the definition of parallel lines according to Lobachevsky. This is not as easy as in the Euclidean case and requires some auxiliary considerations.

Let a be a given line and A a point on this line (Fig. 32). Draw from A a perpendicular AP to a . The line AP divides the plane into two parts, one of which will be called the "right" half plane and the other the "left" half plane. Similarly, a divides the plane into two parts; the one in which A lies will be called the "upper" half plane.

Denote by $\tilde{a}$ the line perpendicular to AP at the point A . From absolute geometry it follows that the lines a and $\tilde{a}$ do not intersect. In view of the Lobachevskian axiom, there are infinitely many lines distinct from $\tilde{a}$ that also do not intersect a . Let α denote the angle formed by the right half line of one of these lines with the half line AP and α_0 the lower bound of the set of such α 's.* Obviously,

$$0 < \alpha_0 < \pi/2$$

Indeed, α_0 is greater than the angle PAM , where M is any point of a situated to the right of P ; hence $\alpha_0 > 0$. Since $\tilde{a}$ is just one of the lines not intersecting a , it follows that $\alpha_0 < \pi/2$.

* That is, α_0 does not exceed *any* of the angles of this set ($\alpha_0 \leq \alpha$), but if α_0 is increased by an arbitrarily small positive quantity ε , then $\alpha_0 + \varepsilon$ exceeds *some* angle of this set.

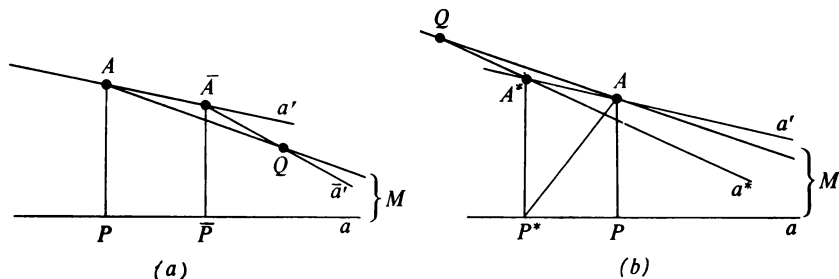


Figure 33

Draw through A a line a' such that its right half line forms the angle α_0 with AP . It is easily seen that a' does not intersect a . In fact, if a and a' intersect, they do so only in the right half plane. Assume that a and a' have a common point R . Choose a point R' on a that lies to the right of R and set $\alpha' = \angle PAR'$. But then $\alpha_0 < \alpha'$ and the lower bound of α happens to be greater than α_0 , contradicting the definitions of α_0 . Denote by a'' the line through A situated symmetrically to a' with respect to AP . The lines a' and a'' form two pairs of vertical angles. Any line passing through A inside that pair of vertical angles where P lies intersects a , while any line through A inside the other pair of vertical angles does not intersect a . As we have just shown, the lines a' and a'' themselves belong to the set of lines which do not intersect a and serve as boundary lines for this set. Of these, a' will be referred to as the *right boundary line* and a'' as the *left boundary line*.

The following important theorem holds.

Theorem II. *Let the lines a and a' be given. If a' is the right boundary line for the set of lines passing through some point of a' and not intersecting a , then a' is the right boundary line for the similar set of lines passing through any other point of this line.*

Proof. Denote by A the point on a' with respect to which the condition of this theorem is fulfilled, and draw from A a perpendicular AP to a .

First we shall prove the theorem for points lying on the right of A . Take any such point $\bar{A}$ and let $\bar{A}P$ be a perpendicular to a (Fig. 33a). It suffices to prove that any ray from $\bar{A}$ situated "below" a' in the right half plane relative to $\bar{A}P$ intersects a . Denote by $\bar{a}$ such a ray. Take any point Q on $\bar{a}'$ and construct the ray AQ . Because the conditions of the theorem are satisfied for A and the ray AQ lies "below" a' in the right half plane relative to AP , this ray must intersect a at some point; call this point M . Since the ray $\bar{a}'$ intersects one of the sides of $\triangle APM$, Pasch's axiom II.4 implies

that it must also intersect one of the remaining two sides of this triangle.* But $\bar{a}'$ cannot have a common point on the side AP , since AP lies in the left half plane with respect to $\overline{AP}$. Accordingly, $\bar{a}'$ has a common point with the side PM (evidently, on the right of P), proving the desired result.

Consider now any point A^* lying on a' on the left of A (Fig. 33b). Let A^*P^* be the perpendicular to a and a^* some ray from A^* lying below a' in the right half plane relative to A^*P^* . We have to show that a^* intersects a . On the complement of a^* take any point Q and join this by a line with A . Since, by our assumption, a' is the boundary for the set of lines through A not intersecting a , it follows that the line QA meets a at some point M lying on the right of P . Note that the ray a^* passes inside the angle AA^*P^* , through its vertex; accordingly, this ray intersects the segment AP^* (by Theorem 11a of Chapter 2). But then, by Pasch's axiom II.4, a^* must intersect either the side AM or the side P^*M of the triangle AP^*M . Since a^* and the line AM have a common point Q lying outside the segment AM , the ray a^* must intersect the side P^*M . Thus the ray a^* and the line a intersect.

If a' happens to be the left boundary line, a similar proof can be given.

Now we define the notion of a parallel line in the Lobachevskian geometry. According to Lobachevsky,

a line a' is said to be parallel to a line a if a' is the boundary line for the set of lines passing through some point of a' and not intersecting a .

Theorem II implies that if a point of a' has the property mentioned in the definition, the same property is possessed by every other point of a' .

We choose one of the two directions on the line a (the direction in Fig. 34 is shown by an arrow) and draw a perpendicular AP to a from a point A of a' . The segment AP forms with a' two adjacent angles, one of which is acute and the other obtuse. If the acute angle is situated on the side of the line AP in the direction of line a , then a' is said to be *parallel to a in the chosen direction* (in diagrams the direction of parallels will be indicated by arrows on both the lines).

* Pasch's axiom II.4 concerns a triangle and a line. This axiom is applicable to a half line (ray) if the vertex of the half line lies outside the triangle and inapplicable otherwise.

In applying Pasch's axiom to a triangle and a half line we should have mentioned that the vertex of the half line lies outside the triangle. However, we shall not mention this each time, thus omitting in this case, and in many others, minor details in the arguments if they are evident enough. Such a scrupulous exposition would fill the book with minor details, which are neither interesting nor important in their essence.

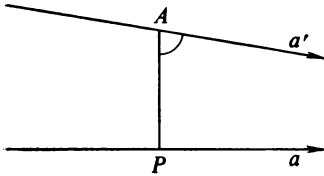


Figure 34

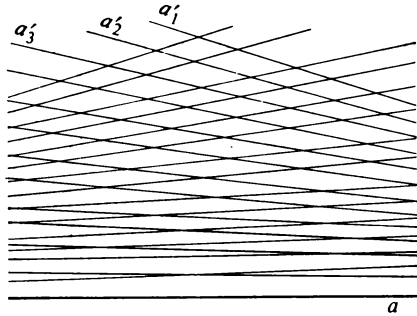


Figure 35

Using the earlier introduced terminology for two sides of the plane with respect to the line AP ("right" and "left"), we can describe the direction of parallelism in a different manner: for the set of lines passing through A and not intersecting a the line a' may be either the right or the left boundary; in the first case a' is said to be *parallel to a on the right* and in the second case a' is *parallel to a on the left*.

Thus, through each point of a plane there pass two lines parallel to the given line in two different directions (see Fig. 35; lines parallel to a on the right are denoted by $a'_1, a'_2, \dots$). In particular, we have the following

Theorem III. *Through each point in the plane there passes just one line parallel to the given line in a chosen direction.*

§ 30. On the basis of the stated definition of parallel lines, we cannot yet speak of two lines being *parallel to each other*. Later we shall establish the relationship of mutual parallelism, that is, if one of two lines is parallel to the second, then the second is parallel to the first. First we shall have to prove some auxiliary propositions.

Lemma I. *Let a and b be any two lines, O a point of b , and let OA be the perpendicular drawn from O to a , where OA forms with b unequal adjacent angles. If x denotes the distance between O and a point of b on the side of the obtuse angle and $y = f(x)$ denotes the length of the perpendicular drawn from this point to a , then $f(x)$ is a continuous function that increases monotonically without bound.*

Proof. Choose two points M and M' on b such that M lies between O and M' (Fig. 36). Draw perpendiculars OA , MP and $M'P'$ to a and set

$$\begin{aligned} OM &= x, & OM' &= x' \\ MP &= y, & M'P' &= y' \end{aligned}$$

where $x' > x$.

Note that the Lobachevskian axiom implies that the sum of the interior angles of the quadrilateral $OMPA$ is less than four right angles; whence, taking into account that the interior angles at the vertices A and P are right angles, we find that $\angle PMM'$ is greater than $\angle AOM$. Consequently, $\angle PMM'$ is obtuse.

On the line $P'M'$ mark off the segment $P'N = PM$. Joining the points M and N , we obtain Saccheri's quadrilateral $PMNP'$ in which the summit angle $\angle PMN$ is acute. Since $\angle PMM'$ is obtuse and $\angle PMN$ acute, the point N lies between P' and M' , that is, $P'M' > PM$. Thus $y' > y$, provided $x' > x$. This shows that $f(x)$ increases monotonically.

Set $\Delta x = x' - x$ and $\Delta y = y' - y$ ($\Delta x > 0, \Delta y > 0$). Obviously, $\Delta x = MM'$ and $\Delta y = NM'$. From the inequality

$$NM' < NM + MM'$$

and noting that NM is smaller than MM' , since the side NM in the triangle NMM' lies opposite the acute angle and MM' opposite the obtuse angle, we find that

$$NM' < 2MM'$$

or $\Delta y < 2\Delta x$.

Similarly, considering the case where M' lies between O and M , we have

$$|\Delta y| < 2|\Delta x|$$

true for all possible positions of M and M' . This means that $\Delta y \rightarrow 0$ as $\Delta x \rightarrow 0$, which proves that $f(x)$ is continuous.

It remains to show that $f(x)$ increases without bound with x . Take a point M'' on b such that $MM' = M'M''$ and draw a perpendicular $M''P''$ to a . For M'' , put $x'' = OM''$ and $y'' = f(x'') = M''P''$. Denoting $h_1 = y' - y$ and $h_2 = y'' - y'$, we have $MP = y$, $M'P' = y + h_1$ and $M''P'' = y + h_1 + h_2$. On the line $P'M'$ mark off

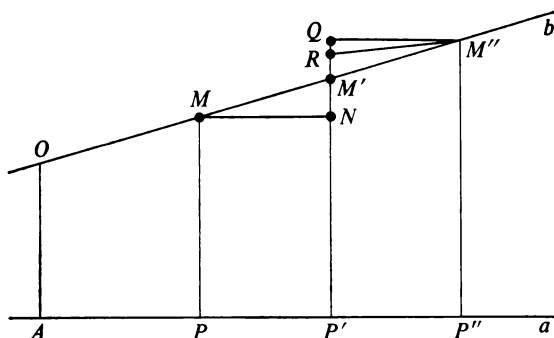


Figure 36

segments $P'N \equiv PM$ and $P'Q \equiv P''M''$ from P' in the direction of M' and the segment $M'R \equiv M'N$ from the point M' . Evidently, $NM' = h_1$, $M'R = h_1$ and $M'Q = h_2$. Noting that the triangles $M'NM$ and $M'RM''$ are equal (they have equal angles included between equal sides), we find that $\angle M'RM'' \equiv \angle M'NM$. But $\angle M'NM$ is adjacent to the acute angle $\angle MNP'$ in Saccheri's quadrilateral. Therefore this angle, and hence the angle $\angle M'RM''$, is obtuse. The angle $\angle M'QM''$ is acute since it is the summit angle of Saccheri's quadrilateral $P'QM''P''$. Comparing $\angle M'QM''$ and $\angle M'RM''$, we find that R lies between M' and Q , that is, $M'Q > M'R$ or $h_2 > h_1$.

From this we have $MP = y$, $M'P' = y + h_1$ and $M''P'' > y + 2h_1$. Thus setting $MM' = M'M'' = s$ and taking the sequence $x_1 = x$, $x_2 = x + s$, $x_3 = x + 2s$, $\dots$, we obtain, successively, $f(x_1) = y$, $f(x_2) = y + h_1$, $f(x_3) > y + 2h_1$, $f(x_4) > y + 3h_1$, and so on. These relations immediately imply that as x increases without bound so does $f(x)$. ■

Note that, although the arguments were heavily based on the Saccheri's quadrilateral in the Lobachevskian system, Lemma I belongs to absolute geometry. The proof of this lemma can be carried out without any trouble for the Euclidean theory of parallels. In doing so, the relations obtained in the concluding portion of the proof are to be replaced respectively by $MP = y$, $M'P' = y + h_1$, $M''P'' = y + 2h_1$, $\dots$, expressing the linearity property of $f(x)$.

An important particular case of Lemma I is the following proposition.

Lemma II. *If x denotes the distance from the vertex of an angle to a point lying on the same side as this angle and $y = f(x)$ denotes the length of the perpendicular drawn from this point to the other side of the angle, then $f(x)$ is a continuous function that increases monotonically without bound.*

Lemmas I and II will be applied several times.

We shall first use Lemma II to prove that the parallelism relationship is mutual.

Theorem IV. *If one of two lines is parallel to the second in a given direction, then the second is also parallel to the first in the same direction.*

Proof. Suppose that a line a is parallel to another line b in a given direction. Then we have to show that b is parallel to a in the same direction.

First we shall show that there is a point equidistant from a and b . This obvious fact follows from Lemma I. Indeed, taking a point P on a , we draw a line PB perpendicular from it to b (Fig. 37). Choose any point M on the segment PB and draw perpendicular MA from it to a .

Set $PB = s$, $PM = x$, $MA = f(x)$ and $\varphi(x) = s - x$. Evidently, $f(x)$ and $\varphi(x)$ denote the distances from M to a and b . By Lem-

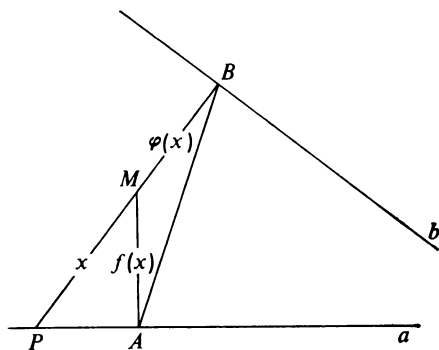


Figure 37

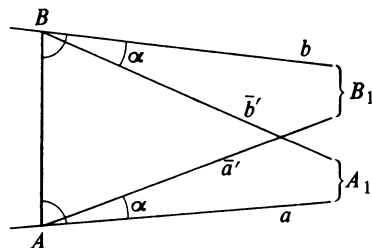


Figure 38

ma I, $f(x)$ is a monotonically increasing function and $\varphi(x)$ is a monotonically decreasing continuous function, as is clear from its expression. Since $f(0) < \varphi(0)$ and $f(s) > \varphi(s)$, there is a unique value of x ($0 < x < s$) such that $f(x) = \varphi(x)$. To this value of x there corresponds a point M equidistant from a and b , that is, $MA = MB$.

For this point M , the line AB is inclined to a and b at equal angles; this line is referred to as the *secant of equal inclination* for lines a and b .

Having established the existence of the secant of equal inclination the mutual parallelism property becomes obvious; however, we shall present a rigorous proof of the theorem.

Since, by the hypothesis, a is parallel to b , we see that a and b do not intersect. Therefore, in order to show that b is parallel to a , we must establish that b is the boundary line for the lines passing through some point of b and not intersecting a . Take B to be such a point (Fig. 38). Let $\bar{b}$ denote the ray with vertex B and pointing in that direction in which a is parallel to b . This ray does not meet a . We have to show that any other ray $\bar{b}'$ having vertex B and inclined to $\bar{b}$ in the direction of a at an arbitrarily small angle α meets a . For a given α , draw through A a ray $\bar{a}'$ which is situated on the same side of a as b and which makes angle α with the direction in which a is parallel. Since a is parallel to b , the ray $\bar{a}'$ must meet b at some point B_1 . On a mark off in the direction of parallelism a segment AA_1 equal to the segment BB_1 . Because AB is the secant of equal inclination for a and b , the triangles BB_1A and AA_1B are equal, whence it follows that the ray with vertex at B which passes through A_1 subtends angle α with b on the side of a , that is, coincides with the ray $\bar{b}'$. This ray intersects a , by construction. Thus a

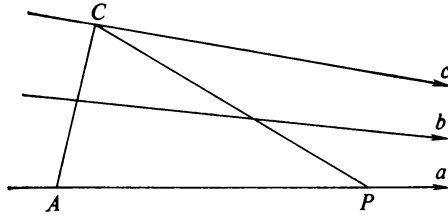


Figure 39

ray passing through B and inclined to $\bar{b}$ at an arbitrarily small angle towards a meets this line; accordingly, b is parallel to a . ■

Consider two lines a and c parallel to each other. The line a divides the plane into two half planes; let Π_{ac} denote the one containing c . In the same way, c divides the plane into two half planes; let Π_{ca} denote the one containing a . Let us agree to call the common portion of the half planes Π_{ac} and Π_{ca} the interior zone of the plane relative to a and c . Suppose that b is a third line parallel to one of the lines a or c , the direction of b being the same as the directions of a and c . It is clear that in this case b intersects neither a nor c . In fact, assuming, for example, that b is parallel to c , we see that b and c cannot intersect, since they are parallel. At the same time, a and b cannot intersect, since otherwise through their common point there would pass two lines parallel to c in the same direction, which is impossible (see Theorem III).

Lemma III. *If under the aforementioned conditions the line b is situated in the interior zone of the plane relative to the lines a and c , then b intersects every segment joining a given point of a with any given point of c .*

Proof. Assume that b is parallel to c . Choose a point A on a , a point C on c and join them by the segment AC . Denote by $\bar{c}$ a ray on c originating from C in the direction of parallelism of a and c (Fig. 39). Let $\bar{c}'$ be the ray originating from C and passing through the interior of the angle between the rays CA and $\bar{c}$; assume further that $\bar{c}'$ lies on that side of the perpendiculars from C to a and b on which the parallel lines are directed. Since, by the hypothesis of the lemma, a and c are parallel, $\bar{c}'$ must intersect a at some point P . Similarly, since b and c are parallel, $\bar{c}'$ must also intersect b , and this point of intersection must lie between C and P because b lies in the interior zone relative to a and c . From this result and Pasch's axiom we conclude that b intersects either the segment AC or the segment AP , but it cannot intersect AP because in this case it does not meet a . Therefore b intersects AC and the lemma is proved. ■

The next theorem establishes the transitive property of parallel lines.

Theorem V. *Two lines parallel to a third line in the same direction are themselves parallel in this direction.*

Proof. Assuming that the lines a and b are parallel in the same direction to the line c , we conclude, as above, that a and b do not intersect (otherwise through their common point there would pass lines parallel to c in the same direction, which is impossible).

To prove that a and b are parallel, we consider two cases (Fig. 40):

- (1) the lines a and b are on the same side of c ;
- (2) the lines a and b are on the opposite sides of c .

In the first case either a or b lies in the interior zone of the plane determined by the other line together with c . Assume, for instance, that b lies in the interior zone relative to a and c .

Take any point A on a and denote by $\bar{a}$ the ray on a which originates at A and points in the direction of parallelism of lines a and c . We must show that $\bar{a}$ is the boundary of the set of rays originating from A and not intersecting b . Assume the contrary, that is, let there exist a ray $\bar{a}'$ which originates from A in the direction of parallelism (that is, which lies on that side of the perpendiculars from A to b and c on which the parallel lines lie) and lies closer to b than the ray $\bar{a}$ but does not meet b . But then, in view of the previous lemma, the ray $\bar{a}'$ cannot meet c either, which contradicts the fact that a and c are parallel, because in this case $\bar{a}$ will not be a boundary for the set of rays originating from A and not intersecting c .

Let us now consider the second case. Assume that a and b lie on the opposite sides of c ; then b and c lie on the same side of a . Draw a ray $\bar{a}'$ through any point A on a so that it is closer to b and c than a and passes on that side of the perpendiculars from A to b and c on which the parallel lines lie. Since a and c are parallel, the ray $\bar{a}'$ intersects c , and since c and b are also parallel, this ray intersects b too. Thus $\bar{a}$ is the boundary line for the set of rays passing through A and not intersecting b ; accordingly, a and b are parallel (in the same direction in which they are parallel to c). ■

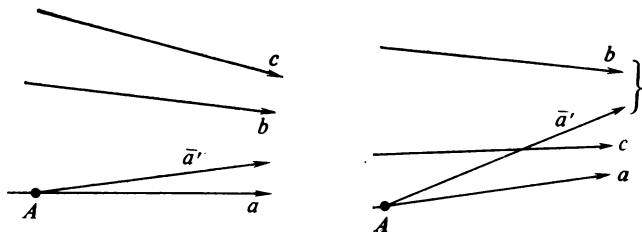


Figure 40

The propositions of the present subsection show that though the definition of parallels in Lobachevskian geometry is far more complicated, the set of lines parallel to a given line in a given direction has the same basic properties as set of parallel lines in Euclidean geometry.

3.2. Peculiarities in the location of parallel and divergent lines

§ 31. If two lines neither intersect nor are parallel, they are said to be *divergent**. Through each point in a plane there are two lines parallel to a given line and an infinite number of divergent lines (Theorem I).

We shall examine some of the properties possessed by parallel and divergent lines, and the results obtained will enable us to clearly distinguish between parallel and divergent lines. First we note the following two theorems.

Theorem VI. *Two lines perpendicular to a third line are divergent.*

The proof of this theorem is straightforward. Indeed, the fact that two lines a and b perpendicular to a third line c at points A and B do not have a common point is known from absolute geometry. But these lines are not parallel either, because through A there pass infinitely many lines not intersecting b and for them a is not a boundary line and hence is not parallel to b . In implicit form this theorem was already established in § 29.

Theorem VII. *If two lines intersected by a third form equal alternate or corresponding angles, they are divergent.*

Proof. This is a generalization of the preceding theorem and the present theorem can be obtained from that theorem. Denote by a and b two given lines and by c a third line intersecting them (Fig. 41). Designating by A and B , respectively, the points of intersection of c with a and b and by O the midpoint of the segment AB , we draw perpendiculars OP and OQ from O to a and b .

In the right triangles OAP and OBQ we have $OA = OB$, by the choice of O , and $\angle OAP = \angle OBQ$, by the assumption of the theorem, whence it follows that $\triangle OAP$ is equal to $\triangle OBQ$. In particular, $\angle BOQ = \angle AOP$, and the segments OP and OQ therefore lie on one line PQ to which a and b are perpendicular. Theorem VI states that these lines are divergent.

We shall now show that any two divergent lines have just one common perpendicular. Thus the presence of only one common perpendicular is characteristic of divergent lines.

* The term "divergent lines" is justified by peculiarities of the location of these lines; see Theorem VIII below.

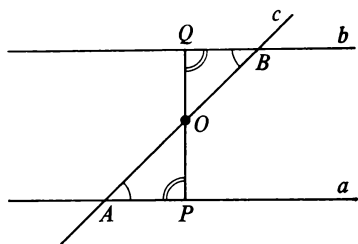


Figure 41

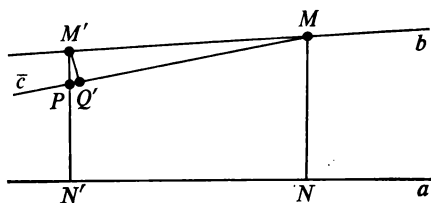


Figure 42

First, it is clear that in Lobachevskian geometry two lines cannot have two common perpendiculars. Indeed, if AB and CD are perpendicular to the lines AC and BD , the quadrilateral $ABCD$ has an angle-sum equal to four right angles, which is impossible since each of the triangles ABC and BCD has an angle-sum less than two right angles. The uniqueness of the common perpendicular is thus immediately established.

Somewhat more complicated is the proof of the existence of the common perpendicular to two divergent lines.

Consider any two divergent lines a and b (Fig. 42), and suppose that MN is a perpendicular from any point M of b to a . If MN is also perpendicular to b , the proof is completed; therefore assume that MN is not perpendicular to b . Choose a positive direction on b so that it makes an obtuse angle with MN . Then, by Lemma I, as M moves in the positive direction, the length MN increases monotonically without bound.

We shall show that after a certain stage the length MN also monotonically increases without bound as M moves in the negative direction.

The line MN divides the plane into two half planes; the one in which the positive half line of b lies will be called positive and the other negative. In the negative half plane, draw through M a ray $\bar{c}$ parallel to a . Since a and b are divergent lines, $\bar{c}$ is closer to a than the negative half line of b is. Therefore, if we take a point M' on b in the negative half plane and draw a perpendicular $M'N'$ to a , this perpendicular will meet $\bar{c}$ at a point P' which lies on the segment $M'N'$. Thus $M'N' > M'P'$. Now if we draw a perpendicular $M'Q'$ to the ray $\bar{c}$, we find that $M'P' > M'Q'$ and consequently $M'N' > M'Q'$. But, according to Lemma II, the distance between a variable point on one side of the angle and the other side increases without bound as the variable point moves farther from the vertex; this implies that as M' moves in the negative direction, $M'Q'$ increases monotonically without bound. Since $M'N' > M'Q'$, $M'N'$ also increases without bound.

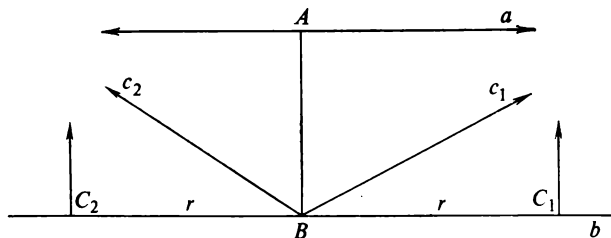


Figure 43

We introduce a linear coordinate system on b by choosing an origin and assuming, for example, that the coordinates increase in the positive direction. Let x denote the coordinate of the variable point M and $y = f(x)$ the length of the perpendicular MN to a . According to the foregoing result, the function $f(x)$ is continuous and everywhere positive and its values increase without bound as x tends to $+\infty$ or $-\infty$. This implies, firstly, that $f(x)$ has a positive minimum $f(x_0)$ and, secondly, that each value greater than $f(x_0)$ is assumed by $f(x)$ for at least two distinct values of the argument. Denote two such values of x by x_1 and x_2 ($x_1 \neq x_2$), so that $f(x_1) = f(x_2)$, by M_1 and M_2 the points with coordinates x_1 and x_2 , and by M_1N_1 and M_2N_2 the perpendiculars from them to a . By the choice of M_1 and M_2 , the quadrilateral $N_1M_1M_2N_2$ is a Saccheri quadrilateral implying its symmetry relative to the perpendicular at the midpoint of the base. This means that a perpendicular M_0N_0 drawn to a from the midpoint M_0 of the segment M_1M_2 is the common perpendicular to a and b , proving the existence of the common perpendicular.

If we now take any point $\bar{M}$ on b and from it draw the perpendicular $\bar{M}\bar{N}$ to a , the angles at the vertices N_0 , M_0 and $\bar{N}$ in the quadrilateral $N_0M_0\bar{M}\bar{N}$ will be right angles and hence the exterior angle at $\bar{M}$ will be an obtuse angle. By Lemma I, this result implies that the values of $f(x)$ increase on the side of $\bar{M}$ that does not contain M_0 . Therefore M_0 is the only point where $f(x)$ attains its maximum.

The foregoing discussion may be summed up in the form of a theorem.

Theorem VIII. *Any two divergent lines have just one common perpendicular on both sides of which these lines move away from each other without bound.*

We note in passing that the projections of all points of one of the two diverging lines onto the second fill only a finite segment on the second line. To see this, let a and b be two divergent lines and AB their common perpendicular (Fig. 43). From B draw lines c_1 and c_2 parallel to a and assume that perpendiculars have been erected at all

points of b . If all these perpendiculars were to meet c_1 and c_2 , it would be necessary to assume the validity of Euclid's fifth postulate (see Proposition 4 of § 8 or Theorem 54 of § 27).

Thus the Lobachevskian axiom implies that the perpendiculars erected at points of b situated at a sufficiently large distance from B do not meet c_1 and c_2 . Denote by r the lower bound of the distances of these perpendiculars from B . Take two points C_1 and C_2 on b such that $C_1B = BC_2 = r$; then evidently the projections of all the points of a fill the interior of the segment C_1C_2 . The perpendiculars at C_1 and C_2 are parallel to a as well as to c_1 and c_2 (this will not be proved here; it is to be found in § 33). If we construct a line that lies symmetrically to a with respect to b , we obtain the diagram in Fig. 44. This is a peculiar "quadrilateral" whose sides and diagonals are parallel in the directions indicated by arrows.

Naturally, this shape has no analogue in Euclidean geometry.

§ 32. We shall now examine the relative position of parallel lines. Let the two lines a and b drawn in Fig. 45 be parallel in a given direction. Taking a variable point M on a , draw a perpendicular MN to b . This perpendicular forms, on the side of parallelism, an acute angle with a . By Lemma I, this implies that the length of MN monotonically increases without bound as M moves in the direction opposite to the direction of parallelism, while in the direction of parallelism it monotonically decreases. We shall show that in the latter case the length MN tends to zero.

Choosing some point A on a , draw perpendicular AB to b . If a positive number ε is given, we must show that for a certain position of M $MN < \varepsilon$. When $AB > \varepsilon$, we take a line b' in the plane and at any point N'_0 of it erect a perpendicular $N'_0M'_0$ whose length is less than ε . Draw through M'_0 a line a' parallel to b' . Suppose that the variable point M' moves along a' in the direction opposite to that of parallelism. The length of the perpendicular $M'N'$ to b' varying continuously increases without bound. Therefore there will be such a location of M' for which the length of $M'N'$ will be equal to AB . Let M' and N' be denoted A' and B' at this stage. We su-

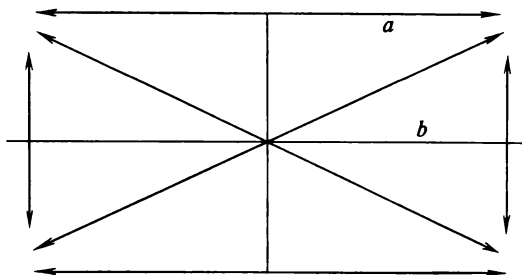


Figure 44

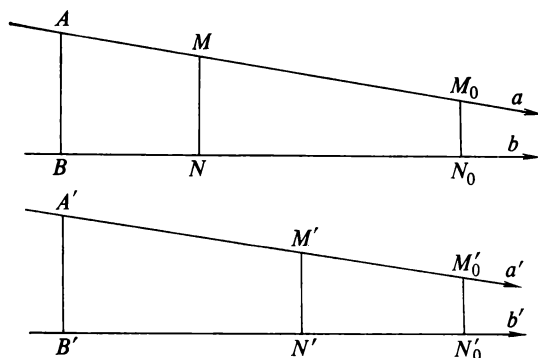


Figure 45

perpose the figure formed by a' and b' in such a way that b' coincides with b , B' with B and the direction of parallelism of a' and b' coincides with that of a and b .*

Since $A'B' = AB$, point A' coincides with A and, since through a given point there passes only one line parallel to the given line in a definite direction, a' coincides with a . Let the point M'_0 of a' assume the position of M_0 on a , and denote by N'_0 the corresponding position of N'_0 . Thus the length of the perpendicular $M'_0N'_0$ is found to be less than any preassigned positive number ε , that is, the lines a and b draw closer without bound in the direction of parallelism.

The foregoing discussion can be summed up in the following theorem.

Theorem IX. *The distance from a variable point of one of the two parallel lines to the second approaches zero if the point moves in the direction of parallelism and it increases without bound if the point moves in the opposite direction.*

The results can be summarized as follows. Two divergent lines always have only one common perpendicular on both sides of which the lines move away from each other without bound ("diverge"); parallel lines moving away from one another in one direction draw asymptotically closer in the other.

[3.3. Lobachevsky's function $\Pi(x)$

§ 33. Consider a given line a and a point A outside it. Through A there pass two lines parallel to a in two different directions; call them u_1 and u_2 . The lines u_1 and u_2 form equal angles with the per-

* When we speak about superposition of figures, we mean the construction of a figure congruent to the given figure. In this sense the notion of motion is completely defined (see § 19).

pendicular AP drawn from A to a . The acute angle formed by either of the lines u_1 or u_2 with the perpendicular AP is called the *angle of parallelism at A relative to the line a* .

Let us prove that the angle of parallelism is completely determined by the distance from A to a .

Let A and A' be two points situated at equal distances from the lines a and a' . Draw through A a line u parallel to a and through A' a line u' parallel to a' . Denote by AP and $A'P'$, respectively, the perpendiculars drawn to a and a' , and by α and α' the angles of parallelism at the points A and A' relative to a and a' . We have to show that $\alpha = \alpha'$. Assume the contrary: one of the angles is less than the other; for example, $\alpha < \alpha'$. Draw through A' a line forming an angle α with the segment $A'P'$ on the side of parallelism of u' and a' . Since u' and a' are parallel, the line drawn must intersect a' at a point Q' (situated in the direction of parallelism from P'). Starting from P , we mark off, in the direction of parallelism, on a a segment PQ equal to the segment $P'Q'$. The triangle PAQ is obviously equal to the triangle $P'A'Q'$ (the sides AP and $A'P'$ are equal by assumption, PQ and $P'Q'$ by construction and the included angles are right angles), therefore $\angle PAQ$ equals α . Consequently, line u coincides with line AQ . But then u must intersect a at Q , which is impossible as the two lines are parallel. This contradiction proves the assertion.

Take any point A not on line a_0 and denote by α the angle of parallelism at A relative to a_0 (Fig. 46). Let x be the length of the perpendicular AA_0 drawn from A to a_0 . By the preceding result, α is completely determined by x ; following Lobachevsky we set

$$\alpha = \Pi(x)$$

The function $\Pi(x)$ plays a fundamental role in non-Euclidean geometry. We shall now examine some of its simple properties.

First we shall show that $\Pi(x)$ is a monotonically decreasing function. To do this, take some point A' on the line AA_0 and let x'

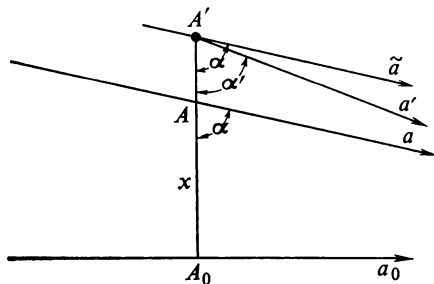


Figure 46

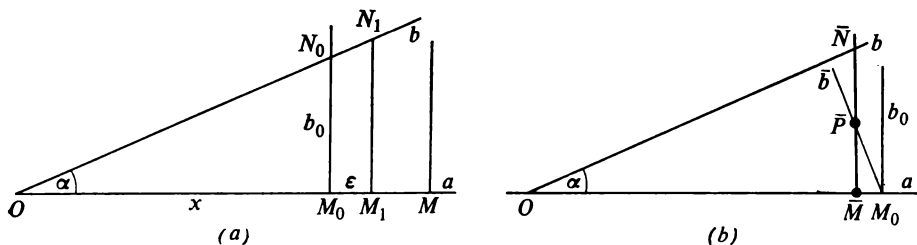


Figure 47

be the length of $A'A_0$. Assume that $x' > x$; we have to show that $\Pi(x') < \Pi(x)$. Draw through A a line a parallel to a_0 and through A' a line $\tilde{a}$ which makes an angle $\alpha = \Pi(x)$ with the segment $A'A_0$ on the side of the parallel lines a and a_0 . The two lines $\tilde{a}$ and a form equal corresponding angles with the line AA' and hence, by Theorem VII, are divergent. This implies that the line a' , which passes through A' and is parallel to a in the same direction as $\tilde{a}$ is parallel to a_0 , is situated on the side of parallelism closer to a than $\tilde{a}$ is. Thus if $\alpha' = \Pi(x')$, then $\alpha' < \alpha$, that is, $\Pi(x') < \Pi(x)$ for $x' > x$.

Next, $\Pi(x)$ assumes all the values between 0 and $\pi/2$. To establish this property, we take any acute angle α and show that it is the angle of parallelism for some segment x . Denote by O the vertex of this angle and by a and b its sides. Lobachevsky's postulate implies that the perpendiculars erected at points of a situated far enough from O do not meet b (see Theorem 54 of Chapter 2 or Proposition 4 of § 8).

Let M be any point at which the perpendicular to a does not meet the inclined line b . Denote by M_0 a point of a such that $OM_0 = x$ is the lower bound of distances OM ; denote further by b_0 the perpendicular to a at M_0 . We shall show that b_0 and b are parallel. To do this, we must first show that b_0 and b do not intersect.

Assuming the contrary, we let N_0 be the common point of b_0 and b (Fig. 47a). Take a point N_1 on b such that N_0 lies between O and N_1 and draw the perpendicular N_1M_1 to a and set $M_0M_1 = \epsilon$. Then, if M is the foot of some perpendicular to a not meeting b , we have $OM > x + \epsilon$, which, however, contradicts the definition of x as the lower bound of OM .

We shall now show that b_0 is the boundary line for the set of lines passing through M_0 and not intersecting b .

Let $\bar{b}$ be any ray through M_0 which lies on the same side of b_0 as O and on the same side of a as the acute angle α (Fig. 47b). Take a point $\bar{P}$ on $\bar{b}$ lying inside α and draw the perpendicular $\bar{P}\bar{M}$ to a . Clearly, $\bar{M}$ will be between O and M_0 and therefore the perpendicular $\bar{P}\bar{M}$ will have a common point $\bar{N}$ with b . Since the ray $\bar{b}$ inter-

sects one of the sides of the triangle $OM\bar{N}$, in fact side $\bar{M}\bar{N}$, from Pasch's axiom it follows that it must also intersect one of the two remaining sides of the same triangle. But $\bar{b}$ does not intersect $\bar{O}\bar{M}$, hence it must have a common point with b . This proves that b and b_0 are parallel and at the same time this proves our assertion. Indeed, for a preassigned acute angle α , we were able to construct a segment $x = OM_0$ such that $\alpha = \Pi(x)$, that is, $\Pi(x)$ assumes all the values between 0 and $\pi/2$.

This also implies the continuity of $\Pi(x)$, since a monotone function assuming together with any two given values all the intermediate values is continuous in the domain of $\Pi(x)$.

Summarizing all the foregoing discussion, we have

Theorem X. *The function $\Pi(x)$ is defined for every positive value of x , decreases monotonically and is continuous; moreover, $\Pi(x) \rightarrow \pi/2$ as $x \rightarrow 0$, while $\Pi(x) \rightarrow 0$ as $x \rightarrow \infty$.*

That $\Pi(x) \rightarrow \pi/2$ as $x \rightarrow 0$ implies that in small regions of space Lobachevskian geometry differs little from Euclidean geometry (since for small x the angle of parallelism is close to a right angle).

The dependence between angular and linear magnitude given by the function $\alpha = \Pi(x)$ gives Lobachevskian geometry its peculiar character. Thus there is no similarity of figures in Lobachevskian geometry. One could foresee this situation. Since the angular and linear magnitudes are connected by an equation, the given angles of a triangle must determine its sides, and triangles with equal corresponding angles are equal. Later (in § 231, Chapter 8) we shall establish this fact precisely and obtain formulas expressing the sides of a triangle in terms of its angles (see also § 64).

One more important peculiarity of non-Euclidean geometry is connected with the choice of a unit for measuring lengths. In Euclidean geometry there are absolute constants of angular quantities, that is, angles whose construction can be described by abstract formulation (independently of the concrete interpretation of geometric objects); even if this construction is arbitrary to some degree, this does not affect the magnitude of the angles. The right angle serves as a good example of this remark. If the right angle is taken as the unit for measuring angles, in the process of measurement there is no need of fixing the "standard" of a right angle with which the remaining angles are compared by superposition, since a right angle can always be found by precise construction.

On other hand, there are no absolute linear constants in Euclidean geometry. In order to express the lengths of all the segments in terms of numbers, we must agree to a choice of a unit of measurement, for which the length of an arbitrary segment may be taken. If someone were to make this choice, he could not describe it and would have to show his standard to compare with it the other segment. In

practice, for the measurement a copy of the standard metre is used, but from the geometric point of view the choice of a standard is not stipulated by anything.

In contrast with this, Lobachevskian geometry contains, apart from the absolute constants of angular magnitude, absolute linear constants, too. For example, the segment x satisfying the equation $\Pi(x) = \pi/4$ is defined in as much as $\Pi(x)$ is. As we have seen, this function is completely defined on the entire positive number axis by the geometric properties of the Lobachevskian plane, that is, by the properties of the manifolds of the geometric figures satisfying axioms of Lobachevskian plane geometry. In § 190, an expression for $\Pi(x)$ will be obtained by means of elementary functions that are well known in mathematical analysis (see also § 59).

3.4. Lines and planes in Lobachevskian space

§ 34. We shall now describe briefly the peculiarities of the relative position of lines and planes in Lobachevskian space.

Before doing so, we shall list the main propositions of absolute geometry which shall be required below.

Without going into the details of the simple consequences of Axioms I.1-I.8 and II.1-II.4, which were mentioned in Chapter 2, we note the following theorems.

(1) *Let there be given a plane α and two lines a and b lying in this plane and passing through a point O of it. If the line c is perpendicular to a and b at O , then it is perpendicular to any line in α passing through O . (In this case the line c is said to be perpendicular to α .)*

(2) *Through a given point in space only one line can be drawn perpendicular to the given plane.*

(3) *Through a given point in space just one plane can pass perpendicular to a given line.*

Two half planes with a common boundary line and not lying in one plane form a *dihedral angle*. The half planes forming a dihedral angle are called *faces* and their common boundary line the *edge*. Any plane perpendicular to the edge of a dihedral angle intersects the faces along two half lines forming the *linear angle of a given dihedral angle*.

(4) *All the linear angles of a given dihedral angle are equal to each other.*

Dihedral angles are *equal* if their linear angles are equal.

A dihedral angle is a *right angle* if its linear angles are also right angles.

Two intersecting planes α and β determine two pairs of vertical dihedral angles. If these angles are right angles, the planes are said to be *perpendicular* to each other.

(5) If a plane α passes through a perpendicular to the plane β , then α is perpendicular to β . (This theorem is clearly a particular case of Theorem 4.)

(6) If α is perpendicular to a line lying in β , then α is perpendicular to β .

(7) Through every line a one can draw a plane α perpendicular to the given plane β and there will be only one such plane if a is not perpendicular to β . (Theorem 7 follows from Theorems 2 and 5.)

The line a' along which planes α and β intersect is called the *projection of a onto β* (if a is not perpendicular to β). According to Theorem 7, each line can be uniquely projected onto any plane not perpendicular to this line.

The above propositions belong to absolute geometry; the propositions given below will be concerned mainly with Lobachevskian geometry.

§ 35. Two nonintersecting lines in Lobachevskian space that lie in one plane are said to be *parallel* or *divergent* if in this plane they are parallel or divergent according to the definitions given in plane geometry.

For later discussion it is important to establish the transitivity property of parallelism, that is, if two lines are parallel to a third in one direction, they are parallel to each other in the same direction. Here we shall be concerned only with the case when the lines are not in one plane; the proposition for the other case has been already established in plane geometry (Theorem V). The transitivity property of parallelism is a direct consequence of the following lemma.

Lemma IV. *The line of intersection of two planes passing through two parallel lines in a given direction is parallel to both these lines in the same direction.*

Proof. Consider two lines a and b lying in a plane γ and parallel to each other in a given direction. Let α and β be planes passing through these lines, respectively, and c be their line of intersection (Fig. 48; the planes α and β are distinct from γ). We have to show that c is parallel to both a and b in the same direction as a and b are.

Let us show, for instance, that a and c are parallel. First, it is clear that a and c do not intersect. Indeed, if these lines were to intersect at some point O , this point would be common to the three planes α , β and γ , in which case a and b would have a common point O , which contradicts the hypothesis.

Take any point C on c and draw perpendicular CA to a . The segment CA forms two adjacent angles with c ; of these we choose the one located on the same side as the parallelism of a and b . Draw a ray $\bar{c}$ through C in the interior of this angle. To show that c and a are parallel, we must show that $\bar{c}$ intersects a .

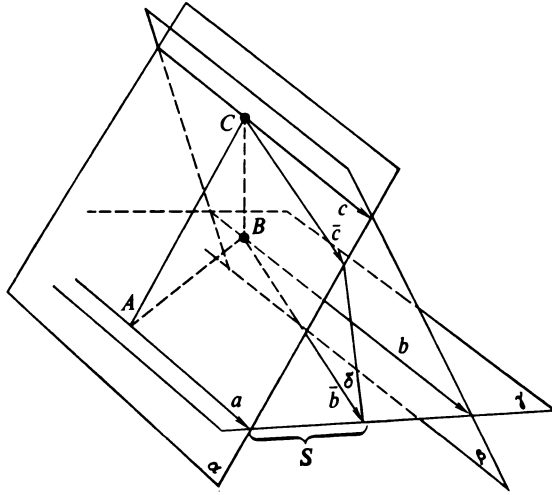


Figure 48

Choosing a point B on b , consider the half plane δ defined by the line CB and the ray $\overline{c}$. This half plane intersects plane γ along some ray $\overline{b}$ which lies in the interior of the angle formed by the segment BA with b in the direction of the parallelism of a and b . Since, by hypothesis, the lines a and b are parallel, the ray $\overline{b}$ intersects a at some point S . Point S is common to the three planes α , γ and δ . Hence $\overline{c}$ must intersect a .

Theorem XI. *Two lines parallel to a third line in one direction are parallel to each other in the same direction.*

Proof. If all the three lines lie in one plane, the theorem has been proved in § 30. Therefore consider three lines a , b and c not in one plane. Suppose that b and c are parallel to a in a given direction. We have to show that b and c are parallel to each other in the same direction as they are parallel to a .

To show this, take any point M on c and pass a plane β through M and b . Let α be the plane in which a and c lie. Since b is not in α , the two planes α and β are distinct. By the previous lemma, line c' , along which α and β intersect, is parallel to both a and b in the same direction as a and b are. According to the hypothesis, c is parallel to a in this direction. But we know that through M there is only one line parallel to a in a given direction. Accordingly, c and c' coincide, that is, c is the line of intersection of α and β , and, by the previous lemma c is parallel to b . ■

This proves the transitivity property of parallelism for space geometry. It should be pointed out that the directions of parallelism of

lines in question must coincide: two lines parallel to a third in opposite directions will never be parallel to each other. The last result may be easily proved if one notes that parallel lines draw closer without bound in the direction of parallelism and diverge without bound in the opposite direction.

In examining the mutual positions of lines and planes we shall restrict ourselves here only to the following three cases.

- (a) A line and a plane have a common point.
- (b) A line is parallel to its projection onto the plane; in this case the line is said to be parallel to the given plane.
- (c) A line and its projection onto the plane diverge; in this case the line is said to diverge from the plane.

We note a theorem which follows immediately from Lemma IV.

Theorem XII. *A line is parallel to a given plane if it is parallel to some line lying in the given plane.*

Noting the discussion in § 32, the reader will easily understand the qualitative differences in the mutual positions of lines that are parallel to a plane and lines that diverge from a plane.

Let us now examine possible cases of the location of planes; three cases will be distinguished.

Case 1. Two planes have a common line.

Let two planes α and β have a common line a . Then in plane α , for example, two lines can be drawn through any of its point parallel to a in various directions. By Theorem XII, these lines will be

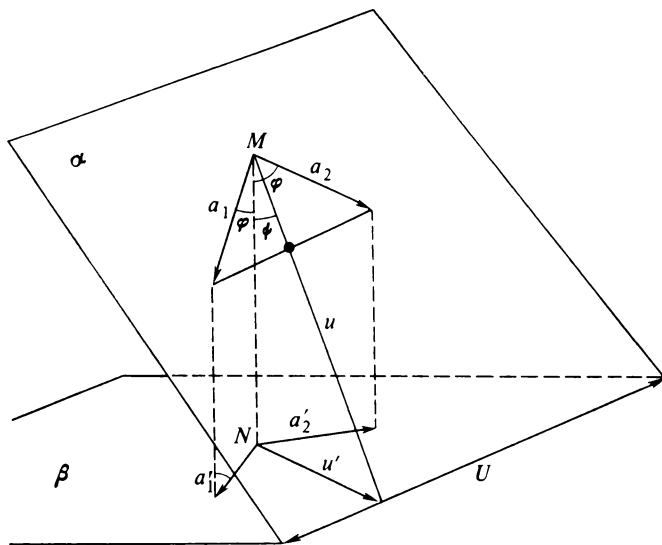


Figure 49

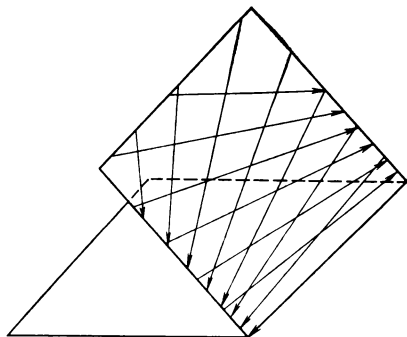


Figure 50

parallel to β . Thus in both of the two intersecting planes, through every point not on the line of intersection there pass two lines parallel to the second plane.

It is not difficult to show that this property is characteristic of intersecting planes. In fact, let planes α and β be given and suppose that in α through any point M there pass two lines a_1 and a_2 parallel to β (Fig. 49). We shall show that α and β intersect. Clearly, a_1 and a_2 form equal acute angles with the perpendicular MN drawn from M to β ; the common magnitude of these angles $\varphi = \Pi(x)$, where x is the length of the perpendicular MN . In α , draw a line u through M such that it passes through the interior of the angle formed by a_1 and a_2 which point in the direction of parallelism. For u we may take, in particular, the bisector of this angle. Denote by ψ the acute angle formed by u with the segment MN . Simple considerations show that $\psi < \varphi$, that is, it is less than the angle of parallelism $\Pi(x)$. Accordingly, u must intersect its projection onto β at some point, which will be the common point of α and β . From this we conclude that α and β have a common line. Thus we obtain the following theorem.

Theorem XIII. *For two planes to intersect it is necessary and sufficient that through any point in one of these planes there pass two lines parallel to the second plane (Fig. 50).*

Case 2. Two planes are so situated that there passes through some point in one of them just one line parallel to the second plane; in this case the planes are said to be *parallel*.

It is clear first of all that under the above condition the two planes cannot have any common point, that is, they do not intersect, since otherwise through each point in one of them there would pass two lines parallel to the second plane.

Next, if of two given planes α and β through point M in α there passes just one line parallel to β or, in other words, parallel to its

projection onto β , then through each point in α there passes just one line parallel to β . Indeed, through each point in α one can draw a line parallel to a in the direction in which the plane is parallel to a' . By Theorem XI, the line thus drawn is parallel to a' and, by Theorem XII, is parallel to β .

In α , there is no other line through the same point parallel to β , since otherwise another parallel line through M would be parallel, by the same reasoning, to β and at the same time would be distinct from a , which is impossible due to our assumption.

Thus plane α is covered by a family of lines parallel to β . It is easy to show that β is similarly covered by a family of lines parallel to α (Fig. 51).

Evidently, both planes draw closer without bound in the direction of the parallelism of the lines belonging to these families.

Case 3. Two planes are so situated that neither of them contains lines parallel to the other; in this case the planes are said to be *divergent*.

Two divergent planes always have a common perpendicular; conversely, planes perpendicular to one line are divergent. Diverging planes move away from each other without bound in all directions from the common perpendicular (hence the name divergent). The proofs of these assertions will not be presented here; the reader can do them himself as exercises.

The above three cases of location of planes can be understood with the aid of the following consideration.

Take a plane α_0 and a point A outside it. From A draw the perpendicular AP to α_0 and through A draw all possible lines parallel to α_0 . All of them are inclined to AP at the same angle $\Pi(AP)$ and hence form a circular cone C with the axis AP (Fig. 52).

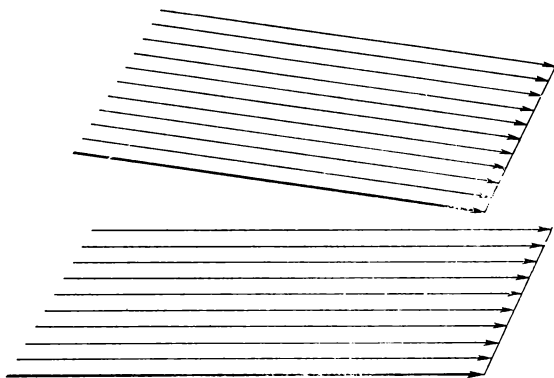


Figure 51

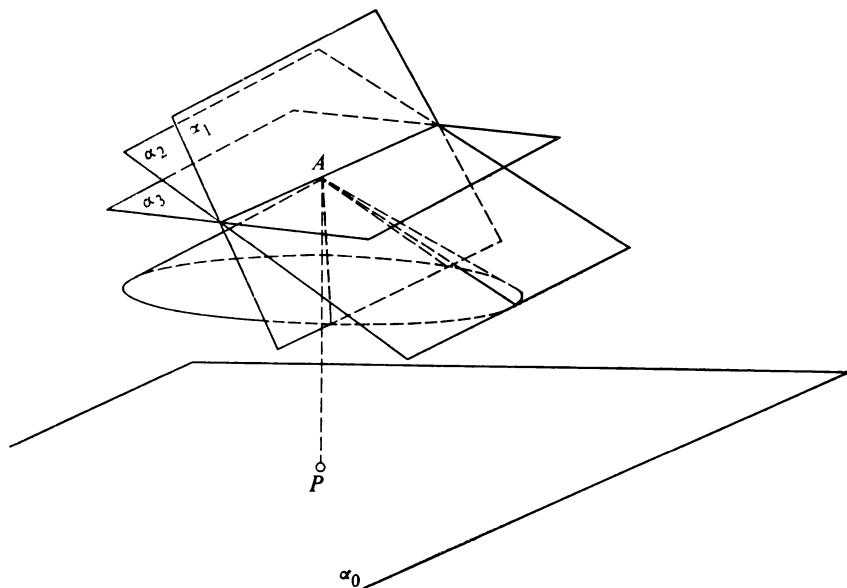


Figure 52

The plane passing through A and intersecting the cone C along two generators contains two lines passing through A and parallel to α_0 (namely, these two generators). Such a plane intersects α_0 (plane α_1 in Fig. 52). The line of intersection with α_0 can be viewed from A through the angle formed by the abovementioned generators of C .

The plane passing through A and tangent to C along a certain generator contains only one line passing through A and parallel to α_0 (the line of contact). This plane is parallel to α_0 (plane α_2 in Fig. 52).

Finally, the plane passing through A and not containing any generator of C does not contain any line parallel to α_0 . This plane diverges from α_0 (plane α_3 in Fig. 52).

We note one more theorem which will be used later.

Theorem XIV. *Given a plane and a line parallel to it, there exists just one plane through this line not intersecting the given plane.*

Proof. Denoting the given line and plane by a and α , respectively, take any point A on a and draw through A all possible lines parallel to α . These lines form a circular cone C with the vertex at A .

If the plane through a does not intersect α , it cannot contain two generators of C and therefore must be tangent to it along the generator a . But through any generator of a circular cone there passes just one tangent plane. ■

With this theorem, we conclude the review begun in § 28 of basic facts about the Lobachevskian theory of parallels. In spite of peculiarities of this theory, in the above exposition one may find many similarities with the Euclidean theory of parallels. In the next section we shall examine a number of important objects of Lobachevskian geometry which have no analogues in Euclidean geometry.

3.5. Equidistants and oricycles

§ 36. In this section we shall deal with certain curves characteristic of non-Euclidean geometry. We shall define them by considering basic forms of motions of the Lobachevskian plane about itself.

At the end of § 19 it was shown that every motion of a configuration consists of a translation along a line and a rotation about a point. Motion was defined as the construction of another configuration congruent to the given configuration. We did not distinguish between the congruent and reflecting configurations. If we consider displacement in the literal sense of this word, that is, if reflection of configurations are omitted, then we may state a much stronger theorem than the one mentioned just now.

In Euclidean planimetry the following theorem holds (in kinematics it is called d'Alembert's theorem).

Each motion of a configuration (or of the entire plane) is either a rotation about a point or a translation along a line.

In other words, the rotation and translation not only combine to give a motion but are the only possible forms of Euclidean motions.

We shall now examine the rotation of the Euclidean plane about a point O . Let k be any circle with its centre at O . When the plane rotates about O , all the points of k are displaced, but remain on the same circle. Thus the circle does not change its position on the plane but "slips" along itself. A curve which retains its position in a motion is said to be *invariant* under this motion.

Concentric circles with the common centre O are clearly invariant under all rotations about O .

If the Euclidean plane is translated along a line u , the invariant curves are the lines parallel to u . In Lobachevskian plane geometry, there are three basic forms of motions:

(1) Rotation about a point: in Lobachevskian plane geometry, like in Euclidean plane geometry, the curves which remain invariant relative to all rotations about a point O are the circles with their centre at O .

(2) Translation along a line: in Lobachevskian plane geometry, curves invariant under all translations along a line u are not lines as in Euclidean planimetry but special curves called *equidistants*.

An equidistant is the locus of points situated on one side of u at equal distances from it. The line u is called the *base* of the equidistant and the distance h its *altitude*; any line may evidently be considered as an equidistant with altitude $h = 0$.

That equidistants are invariant under translations is immediately seen. In fact, when the plane is translated along line u each point of the equidistant with base u is displaced in such a way that its distance from u remains unchanged. This means that this point always remains on the equidistant; thus, on the whole, the equidistant does not change its position.

It is also easy to see that equidistants are curved figures. What is more, the following theorem holds.

Every line has at most two common points with an equidistant.

The proof can be disposed of in two words. Assume that some line has three common points A, B, C with an equidistant, where B lies between A and C . If A', B', C' are projections of A, B, C onto the base, the quadrilaterals $ABB'A'$ and $BCC'B'$ are, according to the definition of an equidistant, Saccheri's quadrilaterals (since segments AA', BB' and CC' are equal). Since Saccheri's acute angle hypothesis holds in Lobachevskian geometry, the angle-sum of ABB' and $B'BC$ is less than two right angles. But since the points A, B, C lie on a line, the sum of these angles must equal two right angles. The contradiction proves the theorem. ■

(3) The third form of motion of the Lobachevskian plane about itself may be called the rotation about a point at infinity.

In order to describe clearly this motion, we require two theorems regarding the "secant of equal inclination".

§ 37. A segment AB whose endpoints lie on lines a and b is called the secant of equal inclination to the lines a and b if it forms with the lines consecutive interior angles.*

Theorem XV. *Given any two parallel lines, through each point on any one of these lines there passes just one secant equally inclined to them.*

Proof. Let a and b be any two lines and S any point at equal distance from a and b (the existence of such a point was established in § 30 in the course of proving Theorem IV). From S draw perpendiculars SP and SQ to these lines. Next, draw the bisector of angle PSQ and call it g . Lines a and b are symmetrical with respect to g . Therefore, if A is any point on a , a point B which is symmetrical to it with respect to g lies on b . The line AB is the desired secant inclined equally to a and b . It is easily seen that through A there is no other secant of equal inclination. Indeed, if the line AB is rota-

* A secant of equal inclination was defined in § 30. For us now it is convenient to use this name for a segment and not a line.

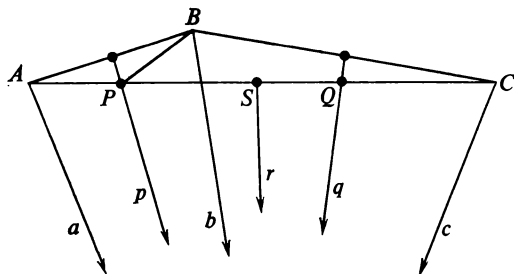


Figure 53

ted about A , then one of the angles formed by it with a and b decreases and the other increases and therefore the rotated line fails to be the secant of equal inclination. ■

Theorem XVI. *In the plane, consider three lines a, b, c parallel to each other in a given direction and passing, respectively, through points A, B, C . If AB is the secant of equal inclination for a and b and BC for b and c , then AC is the secant of equal inclination to a and c .*

Proof. Assume first that b lies between a and c (Fig. 53). Denote by p and q the perpendiculars at the midpoints of the sides AB and BC of the triangle ABC and by P and Q their points of intersection with the side AC .

Since P lies outside the portion of the plane bounded by b and c , $\angle PBC$ is greater than $\angle PCB$. It follows then that the segment PB is less than the segment PC . But $PB = AP$, hence AP is less than PC . A similar argument shows that CQ is less than QA . Therefore the midpoint S of the side AC is between P and Q .

Next note that the line p is parallel to the lines a and b in the direction of their parallelism. Indeed, p can intersect neither a nor b ; if it were to intersect, for example, a , the line b must also pass through this point of intersection because a and b are symmetrical with respect to p . Thus a and b would have a common point, which is impossible as they are parallel. On the other hand, p cannot diverge from either a or b , because they are parallel, a and b draw closer without bound in the direction of parallelism, and as p lies between them, it must draw closer to both of them (to show this fact precisely, it is enough to use Lemma III, § 30). Similarly, q is parallel to b and c . Thus all lines a, b, c, p and q are parallel to each other (in one direction).

Now at S erect a perpendicular r to the side AC . The line r does not intersect both p and q . In fact, if r intersects, for example, p at some point O , this point would be the centre of a circle circumscribing the triangle ABC , in which case q must pass through O . Accordingly, p and q would have a common point O , which is impossible as p and q are parallel. It was noted earlier that S lies between P

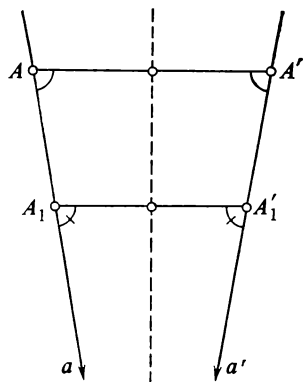


Figure 54

and Q . This together with the abovementioned observation implies that r lies between p and q . Since p and q are parallel in a given direction, r is parallel to them in the same direction. Thus all six lines, a , b , c , p , q and r , are parallel to each other in one direction. For us it is important that r , the perpendicular to AC at its midpoint, is parallel to a and c ; having established this we have almost completed the proof of the theorem. Indeed, this implies that each of the acute angles formed by a and c with AC is $\Pi(AC/2)$, and hence these angles are equal. In other words, AC is the secant of equal inclination to a and c .

Now we examine the case when b does not lie between a and c .

Let AB and BC be secants of equal inclination for their respective lines. Assume that AC is not the secant of equal inclination to a and c . Any one of the three lines is situated between the other two. If a is such a line, for example, then through A we draw the secant of equal inclination AC' to a and c . According to the previous result, BC' is the secant of equal inclination to b and c , and, by hypothesis, BC is the secant of equal inclination for the same lines. This contradicts Theorem XV. ■

§ 38. We shall now explain what is meant by rotation about a point at infinity.

Let there be given a system of all possible lines parallel to each other. We shall imagine these lines meeting in the direction of parallelism at the point at infinity O_∞ . (Speaking of the point at infinity, we only introduce convenient terminology, which in essence does not mean anything except the given system of lines.)

By the rotation about the point at infinity O_∞ we mean a motion of the plane about itself in which any line a of the given system coincides with another line a' of this system (so that a' is parallel to a) and any point A of a is displaced to a point A' of a' such that the

segment AA' is the secant of equal inclination to a and a' (in view of Theorem XV, the position of A' on a' is completely determined by that of A on a); apart from this, it is assumed that the half plane relative to a not containing a' is superposed on the half plane relative to a' containing a . Then

(a) Any other point of line a together with that point of a' to which it is displaced determines a secant of equal inclination to a and a' .

(b) Each line b of the given system while being displaced coincides with line b' of the same system (so that b' is parallel to b), and the corresponding points of b and b' being superposed are the endpoints of secants equally inclined to these lines.

The proof of assertion (a) is evident. Indeed, if A_1 is any point of a and A'_1 the point of a' to which A_1 is displaced, $AA_1 \equiv A'A'_1$ (Fig. 54). Therefore A_1 and A'_1 are situated symmetrically with respect to the perpendicular to the segment AA' at its midpoint (recall that AA' is the secant of equal inclination to a and a'). Since a , a' and A_1 , A'_1 are symmetrically situated with respect to the above perpendicular, $A_1A'_1$ is the secant of equal inclination to a and a' . This proves assertion (a). ■

Somewhat more complicated is the proof of assertion (b). First we introduce the necessary notations. Denote by I the half plane relative to a that does not contain a' , and by II the other half plane; further, denote by I' the half plane relative to a' containing a and by II' its complement. Due to the motion in question of the plane about itself the line a is superposed on a' , the half plane I on I', and the half plane II on II'. Take now any line b in the given system of lines, for example, in I. From A draw a secant equally inclined to a and b and denote by B its endpoint (Fig. 55). We shall locate the point to which B is displaced. From the point A' in I'

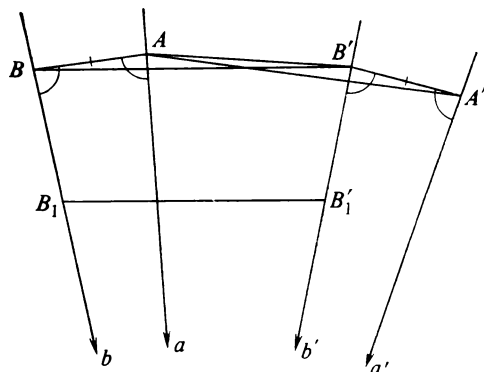


Figure 55

draw a segment equal to AB that makes with a' the same angle as AB makes with a (angles are taken on the side of the direction of parallelism of the lines from the given system); let B' be the endpoint of this segment. The point B' is evidently just that point to which B is displaced. Finally, through B' draw a line b' such that it forms with the segment $A'B'$ the same angle as the one formed by b with the segment AB . Evidently, b' is the line on which b is superposed. Further, it is clear that $A'B'$ is the secant of equal inclination to a' and b' , just as AB is the secant of equal inclination to a and b .

It is also clear that b' is parallel to a' (in the given direction), since b is parallel to a . Accordingly, b' belongs to the given system of lines. Let us show that Theorem XVI implies that BB' is the secant of equal inclination to b and b' . Indeed, since AA' is the secant of equal inclination to a and a' and $A'B'$ is that to a' and b' , Theorem XVI states that AB' is the secant of equal inclination to a and b' . But AB is the secant of equal inclination to a and b ; consequently, the same Theorem XVI implies that BB' is the secant equally inclined to b and b' . Now if B_1 is any point of b and B'_1 the corresponding point of b' under superposition, then $BB_1 \equiv B'B'_1$, which means that $B_1B'_1$ is also the secant of equal inclination to b and b' (see the proof of statement (a)). This proves assertion (b). ■

Now it is easy to understand why the form of motion in question of the plane about itself is called a rotation about a point at infinity. The fact is that if B is any point and B' the corresponding point to which the former is displaced through this motion, then the infinite "triangle" $BB'O_\infty$ (that is, the figure formed by segment BB' and the rays originating from B and B' in the direction of parallelism of the given system of lines) resembles an ordinary isosceles triangle. This resemblance consists in that BB' forms equal angles with the "sides" BO_∞ and $B'O_\infty$.

Thus the point at infinity is in a sense analogous to the centre of ordinary rotation.

The curves which remain invariant under the rotation about a point at infinity were called by Lobachevsky *oricycles*, or *limit circles*.

We shall now indicate a method of constructing these curves, which also will establish their existence.

Consider a system of all possible lines parallel to each other in a given direction. From this system take any line a and a point A on it (Fig. 56). From A draw a secant of equal inclination to a and any other line m of the given system. Denote by M the endpoint of this secant on m . By Theorem XV, the point M is unique.

We now displace line m without taking it out of the given system, that is, retaining its parallelism to a . The point M describes a completely defined curve, which is our oricycle. In other words, an

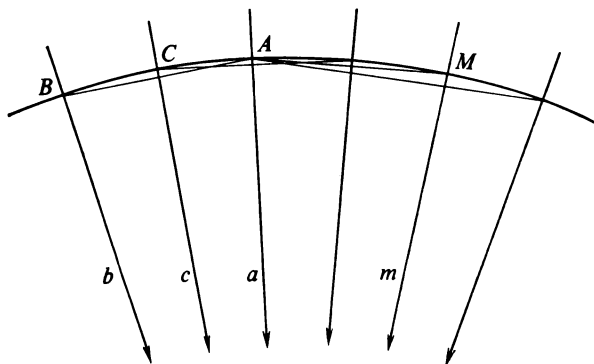


Figure 56

oricycle is a geometrical figure consisting of endpoints of secants of equal inclination drawn from a point A on line a to lines parallel to it in a given direction. The point A itself is assumed to be on the oricycle.

Since the prescribed line a defines the system of lines parallel to it in a given direction, evidently an oricycle is completely determined by prescribing a point A and a directed line a , called its *axis*.

We must show that the oricycle whose construction was just now described is in fact invariant under rotation about the point at infinity O_∞ towards which the axis a is directed.

Let B and C be any points on the oricycle, b and c the lines through these points directed towards O_∞ (that is, parallel to a in the given direction). By the construction of an oricycle, AB is the secant equally inclined to a and b , and AC is the secant equally inclined to a and c . Theorem XVI then implies that BC is the secant equally inclined to b and c . Therefore, if the plane is rotated about O_∞ which takes b to c , then B takes the position of C . Thus under such rotations each point on the oricycle remains on it, as if the oricycle slips along itself. This, in particular, implies that all the points of an oricycle enjoy equal importance and its construction can be started with any point, just as was done with A .

In other words, every line parallel to the axis a of the oricycle in a given direction intersects the oricycle at one point and is also its axis.

For oricycles a theorem analogous to the one established for equidistants is valid.

Any line can have at most two common points with an oricycle.

This, in particular, implies that an oricycle is a curved figure. The proof is left to the reader.

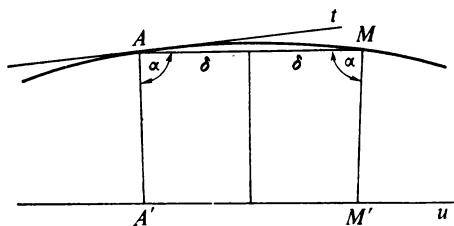


Figure 57

§ 39. Consider an equidistant with the base u . Let A be any point of the equidistant, A' its projection onto the base so that AA' is the altitude of the equidistant (Fig. 57). Draw through A a line t perpendicular to AA' . It may easily be shown that all the points of the equidistant distinct from A lie on the same side of t , namely the side on which u is situated. Indeed, if M is any point on the equidistant and M' its projection onto u , then $AMM'A'$ is a Saccheri quadrilateral and $\angle A'AM$, being the summit angle of this quadrilateral, is acute. Accordingly, M lies on the same side of t as A' . Thus t might be called the supporting line for the given equidistant.* We shall now show that t is tangent as well. Consider the secant AM and let α denote the angle $\angle A'AM$ and 2δ the length of the segment AM . Obviously, the perpendicular at the midpoint of the segment AM and the altitude AA' are divergent lines, as both of them are perpendicular to the base. Therefore angle α is greater than the angle of parallelism for segment δ , that is

$$\alpha > \Pi(\delta)$$

On the other hand, since α is acute, we have

$$\Pi(\delta) < \alpha < \pi/2$$

If the point M approaches A along the equidistant, then $\delta \rightarrow 0$, and, by Theorem X, $\lim_{\delta \rightarrow 0} \Pi(\delta) = \pi/2$. Hence

$$\lim_{M \rightarrow A} \alpha = \frac{\pi}{2}$$

Thus we have shown that as $M \rightarrow A$, the secant AM assumes a limiting position t .

The above result can be expressed in this form: each altitude of the equidistant is a normal to it. The previous result also implies that at each point the equidistant is concave to the base.

We shall now carry out a similar investigation for the oricycle.

* A line is called supporting for the given curve if it contains at least one point of the curve and on one side of this line there are no points of this curve.

Consider an oricycle defined by a point A and axis a (Fig. 58). The direction of axis a , like any other axis of the oricycle, is assumed to be positive in which this axis is parallel to other axes of the oricycle. Draw through A a line t perpendicular to the axis a . It is easy to see that all points of the oricycle, distinct from A , lie on one side of t ; namely, on the side of the positive direction of the axis a . Indeed, let M be any point on the oricycle and m an axis through M . Denote by α the angle formed by the segment AM with the positive direction of the axis a , and by 2δ the length of the segment AM . Since, by the definition of an oricycle, AM is the secant of equal inclination to lines a and m , the perpendicular to the segment AM erected at its midpoint is parallel to both a and m . Therefore α is the angle of parallelism for the segment δ :

$$\alpha = \Pi(\delta)$$

This implies, first of all, that α is acute. Accordingly, any point M of the oricycle does lie on that side of t which is the positive direction of a . In other words, t might be called the supporting line for an oricycle. But t is tangent as well; this follows from

$$\lim_{\delta \rightarrow 0} \Pi(\delta) = \frac{\pi}{2}$$

which implies that as $M \rightarrow A$ the secant AM attains the limiting position t .

The last result can also be formulated by saying that each axis of the oricycle is a normal to it. It is also clear from the foregoing discussion that an oricycle is concave in the positive direction of its axis. We note some properties common to a circle, an oricycle and an equidistant:

(i) Each of these curves is symmetric with respect to any of its normals. Therefore normals to a circle and an equidistant as well as to an oricycle will sometimes be referred to as axes.

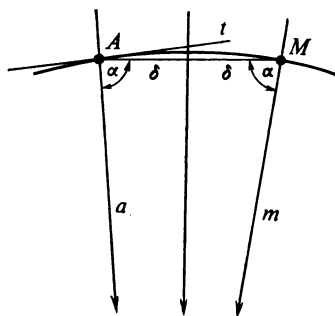


Figure 58

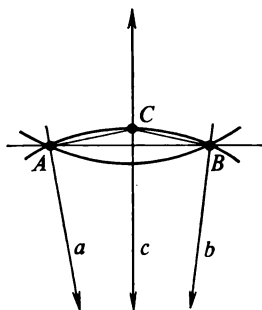


Figure 59

(ii) The chords of these curves are secants of equal inclination to normals passing through their endpoints. Combining a circle, an oricycle and an equidistant we may characterize the family of their normals as follows: all the normals to a circle converge at one point; all the normals to an oricycle are parallel to each other in a given direction (as we say, they converge at infinity); all the normals to an equidistant are perpendicular to a line, and hence diverge.

In Euclidean geometry the set of lines through a given point of a set of parallel lines is called a *pencil*. Extending this notion to Lobachevskian geometry, we shall mean by a pencil a set of lines through a given point or a set of lines parallel to each other in a given direction or a set of lines perpendicular to a given line. For these three cases the pencil will be respectively called *elliptic*, *parabolic* and *hyperbolic*. In view of the aforementioned observation, we may say that circles, oricycles and equidistants are orthogonal trajectories of elliptic, parabolic and hyperbolic pencils, respectively.

§ 40. It is important to observe that whereas circles differ from one another on account of radii and equidistants on account of altitudes, all oricycles are congruent to each other.

Indeed, as we have seen, an oricycle is completely determined once a point on it and an axis through this point are given. Therefore, if the plane is moved in a manner such that the point and the axis through it of one oricycle coincide respectively with the point and the axis of any other given oricycle, then both oricycles coincide (the properties of motion to be used in this connection are ensured by Theorem C of § 19).

We shall prove the following

Theorem XVII. *Given two points A and B in the plane, just two oricycles pass through them and both are symmetric with respect to the line AB.*

Proof. Draw through the midpoint of the segment AB a line *c* perpendicular to AB (Fig. 59), and on this line choose a definite di-

rection. Furthermore, through A and B draw lines a and b parallel to c in the chosen direction. Let AC be the secant equally inclined to a and c . Then, by Theorem XVI, secant BC is equally inclined to b and c . The oricycle defined by C and axis c evidently passes through A and B .

If the opposite direction is chosen on c and the whole construction is repeated, another oricycle is obtained which is symmetrical to the first with respect to AB .

We shall now show that there are no more oricycles through A and B . For this, assume that there is another oricycle L with chord AB , and denote by a and b the axes passing through the endpoints of this chord. Lines a and b must be parallel and equally inclined to the segment AB . Therefore the perpendicular c at the midpoint of AB is parallel to both a and b . But then a and b are completely determined by the direction of parallelism to c , and, accordingly, there are only two possible ways in which a and b can be situated, both examined earlier. Therefore L necessarily coincides with one of the two oricycles constructed in the first part of the proof. ■

This theorem can be alternatively formulated as

Theorem XVIII. *Arcs of oricycles bounded by congruent chords are congruent.*

3.6. Equidistant surfaces and horospheres

§ 41. The space analogue of a circle is a sphere. There are surfaces which are natural analogues of an equidistant and an oricycle. They are respectively called an *equidistant surface* and a *horosphere*. An equidistant surface is a geometrical figure consisting of points situated on one side of a given plane σ and at equal distances from it. The plane σ is called the *base* of the equidistant surface and the perpendicular drawn from any point of the surface to the base is called its *altitude*. This definition is similar to that of an equidistant. [Similarly, the definition of a horosphere is completely analogous to that of an oricycle.

In order to formulate this definition, take in space a line a passing through a given point A . Choose any one of the two directions on a and call it the positive direction. Let m be any other line in space parallel to a in the positive direction. According to Theorem XV, through A there passes only one secant equally inclined to the lines a and m . Denote by M the endpoint of this secant lying on m . If we displace m without disturbing its parallelism to a in the positive direction, the corresponding points M produced a surface, which is called a horosphere.

In other words, a horosphere is the locus of endpoints of equally inclined secants drawn from a given point A of a line to all lines in

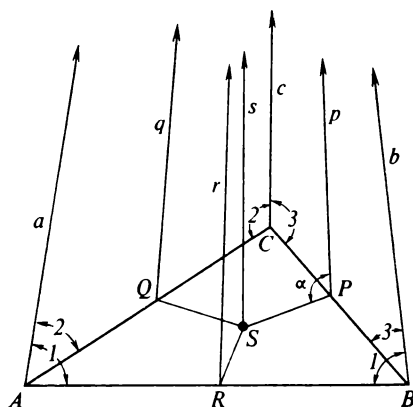


Figure 60

space parallel to the given line in a chosen direction. Point A itself is assumed to be on the horosphere.

Since the prescribed line a determines a system of lines in space which are parallel to it in a chosen direction, a horosphere is completely defined by prescribing a point A and a directed line a , called the *axis of the horosphere*.

It is important to establish that A does not enjoy any privilege among points of the horosphere, that is, the horosphere can be constructed, according to its definition, by starting with any of its points. To do this, we must show that given two lines parallel to the axis of the horosphere in the chosen direction, the segment joining their points of intersection with the horosphere is a secant of equal inclination to them.

Everything, evidently, depends on proving the following

Theorem XIX. *Given three mutually parallel lines a, b, c in space through points A, B, C , respectively, if secant AB is equally inclined to a and b , secant BC is equally inclined to b and c , then secant AC is equally inclined to a and c .*

Proof. It will be assumed that the lines a, b, c do not lie in one plane, the case where all of them lie in one plane was treated earlier in Theorem XVI.

Denote by P, Q, R the midpoints of the sides of triangle ABC opposite the vertices A, B, C and through P, Q, R draw lines p, q, r parallel to a, b, c (Fig. 60), and hence parallel to each other.

It is easy to see that the projections of p, q, r onto the plane ABC meet at one point. Indeed, since p and r are parallel, at least one of them, p , say is not perpendicular to the plane ABC . Denote by α the acute angle formed by this line with the plane ABC and mark

off segment PS on the side of this angle lying in the plane ABC such that

$$\Pi(PS) = \alpha$$

Let s be the perpendicular to the plane ABC at the point S . By its construction, the line s is parallel to p , but since p and q are parallel to each other, it follows that s is parallel to q and r as well. This implies that QS and RS are projections of q and r , that is, all three projections meet at S .

Note further that r is perpendicular to AB , since AB is the secant of equal inclination to a and b ; arguing similarly, we find that p is perpendicular to BC . But the segments PS and RS are perpendicular respectively to BC and AB ; consequently, S is the centre of a circle circumscribing the triangle ABC . In view of this, line AC is perpendicular to QS , that is, to the projection of q ; hence, it is perpendicular to q .

Thus q is a perpendicular to the midpoint of AC . Since q is parallel to a and c , it immediately follows that AC is the secant of equal inclination to a and c . ■

By the same token, it has been demonstrated that the aforementioned construction of a horosphere can be carried out starting with any of its points.

The result can alternatively be expressed thus: each line parallel to the axis of a horosphere in the positive direction intersects the horosphere at one point and is also its axis.

§ 42. We note some common properties of spheres, horospheres and equidistant surfaces.

Consider first the sphere. The properties to be mentioned now do not depend on the fact whether Lobachevskian space or Euclidean space is under consideration. The reader is of course well acquainted with these properties and they are given here only to compare with similar properties of equidistant surfaces and horospheres. Take any point A on a sphere and let a denote the diameter terminating at A . Every plane passing through diameter a intersects the sphere along great circle. Obviously, all great circles obtained by such sections have at the common point A tangents perpendicular to the same line a . These tangents then lie in one plane, called the *tangent plane to the sphere at A* . The diameter a is perpendicular to the tangent plane and hence is a normal. Therefore we may say that all the normals to the sphere meet at one point (the centre of the sphere).

Consider now the equidistant surface. Take any point A on it, and let a be its altitude through A . Obviously, every plane α passing through altitude a intersects the equidistant surface along an equidistant. The base of this equidistant is the line along which plane α intersects the base of the equidistant surface, while the altitude is the same as that of the equidistant surface. From investigations of

§ 39 it follows that tangents at the common point A to all equidistants obtained by such sections are perpendicular to one line a . These tangents then are situated in one plane, called the *tangent plane to the equidistant surface at A* . The altitude a is perpendicular to the tangent plane and hence is a normal. But since altitudes are perpendicular to the base σ , we may say that all normals to an equidistant surface are perpendicular to one plane.

Consider, finally, the horosphere. Let A be any point on the horosphere and a its axis passing through A . Obviously, every plane α passing through the axis a intersects the horosphere along an oricycle, and the axis a of the horosphere is also the axis of the oricycle. The results of § 39 imply that tangents at the common point A to all oricycles obtained by such sections are perpendicular to one line a . Accordingly, these tangents lie in one plane, called the *tangent plane to the horosphere at A* . The axis a is perpendicular to the tangent plane and hence is a normal. But since all the axes are parallel to each other in one direction by the definition of a horosphere we may say that all the normals to a horosphere form a system of mutually parallel lines.

§ 43. A system of lines in space is called a *bundle* if any two lines of this system lie in one plane. Lines constituting a bundle will be called rays.

Let a and b be any two rays. Since, by the definition of a bundle; a and b lie in one plane, the two lines may be located in three ways:

- (1) a and b intersect at some point,
- (2) a and b are parallel in some direction,
- (3) a and b are divergent.

We will now examine each of the three cases.

(1) Assume that a and b intersect at a point O . Let c be any third ray outside the plane of a and b . Denote by α the plane containing rays a and c and by β the plane containing rays b and c . Both planes go through O ; hence c , being the line of intersection of α and β , must also pass through O .

Take now any ray d in the plane of a and b . Since a and c pass through O and d does not lie in the plane of a and c , arguing similarly we find that d also passes through O . Accordingly, all the rays pass through one point. Such a bundle is known as an *elliptic bundle* and the point where all these rays meet the *centre* of the bundle.

(2) Assume that a and b are parallel in a given direction. Let c be any third ray not in the plane of a and b . Let α be the plane containing a and c , and β the plane containing b and c . Since α and β pass through two parallel lines a and b , Lemma IV § 35 implies that c , being their line of intersection, is parallel to a and b in the same direction. Choose any ray d in the plane of a and b . Since a and c are parallel and d is not in the plane of a and c , we find, by

arguments similar to those used in the previous case, that d is parallel to a and c . Thus all the rays of the bundle are parallel to each other in a given direction. Such a bundle is called a *parabolic* bundle.

(3) Finally, consider the case when a and b diverge. Then there is a plane σ perpendicular to these rays. Denote by c any third ray of the bundle which does not lie in the plane of a and b . Let the plane α contain rays a and b , and the plane β rays b and c . Both α and β are perpendicular to plane σ , as the first contains line a perpendicular to σ , while the second contains line b perpendicular to σ . But then c , the line of intersection of α and β , must be perpendicular to σ . Choose any ray d in the plane of a and b . Since a and c are perpendicular to σ and d is not in the plane of a and c , we find, by arguments similar to the preceding, that d is perpendicular to σ . In this case all the rays of the bundle are perpendicular to one plane. This bundle is called a *hyperbolic* bundle and the plane perpendicular to its rays is called the *base* of the bundle.

Summarizing, we have the following proposition.

Spheres, horospheres and equidistant surfaces all have one common property, namely, that normals to each of these surfaces form a bundle. Furthermore, the bundle formed by normals to a sphere is elliptic, to a horosphere is parabolic and to an equidistant surface is hyperbolic.

§ 44. One more common property of spheres, horospheres and equidistant surfaces which will prove to be important later is the following: each of these surfaces is a surface of revolution about any of its normals.

The proof of this fact, which is quite obvious, will be given only for the case of a horosphere.

Let Σ be any horosphere, A a point on it and a the normal through A . We shall take into account all possible rotations of space about the line a (see § 19). We have to show that under rotations all points of Σ , being displaced, remain on Σ , or, using the language of § 36, Σ is invariant under rotations about the line a . To do this, take any point M on Σ and let M' be the point to which M is displaced when the space rotates about a . Further, let m be the normal to the horosphere through M and m' the line coinciding with m under the rotation; evidently, m' passes through M' . The known properties of a horosphere imply that m is parallel to a and the segment AM is the secant of equal inclination to these two lines. But the figure formed by a , m' and the segment AM' is congruent to the figure formed by a , m and AM . Hence m' is parallel to a , and AM' is the secant of equal inclination to a and m' . This means that M' is on Σ .

The proof of this proposition for the case of a sphere and an equidistant surface is also quite easy.

3.7. Elementary geometry on surfaces in Lobachevskian space

§ 45. For a long time two geometric systems on two-dimensional manifolds of Euclidean space have been well known: geometry in a plane (plane geometry) and geometry on a sphere. In the development of these geometrical systems a fundamental role is played by the fact that both the plane and the sphere can move about themselves without being deformed.

In the light of the definition given in § 19, the precise meaning of this statement may be expressed thus: a surface admits movement about itself if there are congruent transformations under which points of the surface remain on it.

If we imagine, for example, a sphere in the form of a smooth wooden model covered tightly by a thin strong cover, then the motion of this cover about the fixed model will enable us to visualise this notion.

A plane and a sphere are not the only surfaces of Euclidean space which can move about themselves, but they distinguish themselves among other such surfaces by the large degree of freedom enjoyed by motions admissible on them.

Any surface of revolution also admits motion about itself, but this property differs from the corresponding property of a sphere or a plane in the freedom of choice of motions. In order to make this difference clear, let us compare, for example, a sphere, a right circular cylinder and an ellipsoid of revolution.

The only motions of an ellipsoid about itself are the rotations about its axis. Every point of an ellipsoid is displaced along a definite trajectory such that for any two such points there is, generally speaking, no motion which would make them coincide.

A right circular cylinder has, apart from rotation, a motion parallel to the axis; combining these two motions any point of the cylinder may obviously be made to coincide with any other point.

A collection of motions which can be performed on a surface is said to be *transitive* if any two points of the surface can coincide through a certain motion. A right circular cylinder then has a transitive collection of motions; in contrast, the collection of motions of an ellipsoid of revolution is nontransitive.

It is easy to see that, like a right circular cylinder, the collection of motions of a sphere is transitive. However, there is an important difference here. To find this difference, the linear element of a surface will be brought into play. A *linear element* is a point together with a given direction, indicated by an arrow, originating at this point in the tangent plane. Linear elements are coincident if their points coincide and their arrows point in the same direction.

On a right circular cylinder, take two linear elements whose points can occupy any position and suppose that the direction of one of them is perpendicular to the axis of the cylinder while that of the second is parallel to the axis. Through motions we can make the points of these linear elements coincide, but the linear elements themselves cannot be made to coincide.

On the other hand, for any two linear elements of a sphere there is always a motion which makes them coincide. By rotating a sphere about an axis first the points of the linear elements can be made to coincide, and then by rotating about the axis on which the coincident points lie the directions can also be made to coincide.

A collection of motions performed on a surface will be called *transitive relative to linear elements* if any pair of linear elements on this surface can be made to coincide.

We may thus say that an ellipsoid of revolution, for example, has a nontransitive collection of motions, while the collection of motions of a right circular cylinder and a sphere is transitive; for the latter, it is even transitive relative to linear elements. Also transitive relative to linear elements is the collection of motions of a plane.

In the basic propositions of plane geometry concerning comparison of geometric quantities, considerable use is made of the possibility of moving the plane figures fairly freely. For example, in defining the length of a rectilinear segment AB , a segment starting from A with a preassigned length equal to unity is marked off on it as many times as possible without crossing B . In this way the length of AB is determined to within integral units. Similarly, by determining how many times half of a unit of measurement is placed on AB , the length of AB is found to within one-half, and so on within any accuracy (see § 20). Thus the measurement is based on the possibility of moving a segment in such a manner that its initial point coincides with a predetermined point and the segment itself is put on an arbitrary given line passing through this point. In other words, use is made of the fact that the collection of motions of a plane is transitive relative to its linear elements.

In spherical geometry, the role played by lines in plane geometry is taken over by the great circles of a sphere. This is due to the following three factors:

(1) Among all curves joining any two points on a sphere the shortest is the arc of a great circle.

(2) Through any two points on a sphere, not being diametrically opposite, there passes only one great circle.

(3) A great circle is determined by any of its linear elements. (A linear element of a curve is the linear element whose point lies on this curve and the arrow points along the tangent to this curve.)

In developing spherical geometry, we can carry out measurements of geometric quantities on a sphere by considering them as objects

of space geometry. For example, we can define the length of an arc of a great circle by taking it equal to the exact upper bound of the lengths of the inscribed polygonal lines whose vertices lie on this arc and whose parts are rectilinear segments. The same procedure is adopted in defining the length of an arc of any curve in space.

But spherical geometry can be developed without using geometric objects not situated on a sphere (like rectilinear parts of the inscribed polygonal lines). This can be achieved in the same way as in plane geometry. For example, in order to extend to spherical geometry the above process of measuring a rectilinear segment, we must start by choosing a unit of length. Suppose that the length of an arc of a great circle is taken as the unit (for the sake of clarity, the reader is recommended to consider this arc as being smaller than the size of the sphere). If some arc AB of a great circle needs to be measured, by displacing this unit of measurement on the sphere we put it on AB , starting from A , as many times as possible without crossing B . Thus AB is measured to within integral units. Similarly, the length of arc AB is found to within one-half of the unit depending on how many times the half of the unit is placed on the arc; proceeding in this way we find the length to within any desired accuracy. Evidently, this process makes considerable use of the transitivity of the collection of motions of a sphere relative to its linear elements.

Other geometric objects (angles, areas) are similarly measured by means of superposition on the given spherical figure of a prechosen unit or its parts. In this process there is no need to use objects of space not belonging to a sphere.

Geometry on any surface may be similarly examined. The role of lines in this general case is taken over by geodesics. A *geodesic* can be defined as a curve each sufficiently small arc AB of which is shorter than any other arc on the surface with the same endpoints A and B . On a sphere of radius R , for example, the great circles are geodesics, since the arc of a great circle whose length is less than πR is shorter than any other arc on the sphere with the same endpoints.

Under certain conditions of an analytic nature to be satisfied by the surface, it may be shown that a geodesic is defined by its linear element, that is, by a point and a direction, just like a line in the plane.

Suppose now that in some way the collection of all the geodesics on a surface has been found. Then, if geometry is being developed on the surface, a natural question may be asked as to whether the lengths of segments of geodesics can be compared in the same way as in plane geometry or in spherical geometry. For this, obviously, by moving the surface about itself, we must be able to displace the arc of a geodesic chosen as the unit of measurement in a way such that its initial point coincides with any point and that it assumes any given

direction. In those cases when the lengths of geodesics may be compared by the superposition of the unit of measurement we shall say that the surface admits of *'elementary geometry'*. Obviously, for elementary geometry to be defined on a surface, it is necessary that the collection of its motions about itself be transitive relative to its linear elements.

It may be shown that the only surfaces in Euclidean space having a collection of motions relative to linear elements are a plane and a sphere. This implies that in Euclidean space two-dimensional elementary geometry is possible only on a plane (plane geometry) and on a sphere (spherical geometry).

In Lobachevskian space, there are two more surfaces, besides a plane and a sphere, on which elementary geometry can be developed. They are an equidistant surface and a horosphere.

That these surfaces indeed allow motions about themselves whose collection is transitive relative to its linear elements was in fact established in § 44, where it was shown that each of these surfaces is a surface of revolution about any of its normals. Indeed, if two linear elements are taken on an equidistant surface or on a horosphere, then by rotating the surface about some normal the points of these linear elements coincide and then by rotation about the normal passing through the coincident point the linear elements themselves coincide.

Thus we may state that in Lobachevskian space elementary geometry may be defined, apart from the plane, on a sphere, an equidistant surface and a horosphere.

Geometry on a sphere in Lobachevskian space does not differ from this geometry in Euclidean space, hence this geometry (spherical) will not be treated here. On the other hand, we will try to give a brief account of geometry on an equidistant surface and a full account of geometry on a horosphere.

§ 46. Consider any equidistant surface Σ whose base is the plane σ . In line with the general ideas presented in § 45, for geometry on Σ we must take as lines the geodesics of this surface. The geodesics of Σ are equidistants obtained by intersecting this surface by planes perpendicular to the plane (that this is in fact so is left without proof for the time being). Therefore the role of lines on Σ is played by these equidistants.

Our aim is to discover a system of propositions which enable us to deduce in a logical way all the properties of mutual relationships between points and equidistants of Σ , that is, to give an axiomatic justification of the geometry of equidistant surfaces. As we shall now demonstrate, this geometry can be substantiated by axioms of Hilbert's first four groups and the Lobachevskian parallel axiom. (It should be noted that, since we are discussing two-dimensional geometry, Axioms I.4-I.8, having a space character, must be omitted

from Hilbert's axioms; therefore we need only use Axioms I.1-I.3, the axioms of Groups II-IV and the parallel axiom.)

To obtain the desired result in as simple a way as possible, we project the points and equidistants of Σ on the plane σ . The projections will be points and lines, respectively. We say that images Φ and Φ' , where Φ is on Σ and Φ' on σ , correspond to each other if Φ' is the projection of Φ . Obviously, points and equidistants on Σ have the same relationship in regard to their location and order as the corresponding points and lines in σ . Therefore Axioms I.1-I.3 and the axioms of Group II are valid for geometry on Σ , since they are valid in plane geometry.

Further, two images on Σ will be said to be *congruent* if they can be made to coincide through a certain motion of Σ about itself and also if they are symmetrical about a certain plane. We correlate all possible displacements of Σ and σ with each other, so that Σ and σ form a solid body in space. Then every movement of σ through which two of its images Φ' and Ψ' coincide causes a movement of Σ such that images Φ and Ψ , corresponding to Φ' and Ψ' , on it coincide. In other words, the images on Σ have the same congruence relationship as their corresponding images on σ . From this we can conclude that the axioms of Group III (axioms of congruence) are valid for geometry on Σ because they are valid for plane geometry.

In the same way it can be demonstrated that the axioms of Group IV (axioms of continuity) are valid for the geometry on surface Σ .

Consider now on Σ any equidistant a and a point A outside this equidistant. Projecting a and A on σ , we obtain a line a' and point A' . Suppose that through A' in σ a line b' is drawn; b' is the projection of some equidistant b lying on Σ and passing through A . If b' does not meet a' , the equidistant b also has no common point with equidistant a . But Lobachevskian geometry is valid on σ and therefore through A' there pass infinitely many lines not intersecting a' . Hence through A on Σ there pass infinitely many equidistants not intersecting equidistant a . This means that the Lobachevskian parallel axiom is realized on Σ .

Thus all the axioms of absolute geometry and the Lobachevskian axiom are valid on Σ . Therefore all the theorems of non-Euclidean plane geometry are valid for points and equidistants on Σ .

Therefore, elementary geometry on an equidistant surface is Lobachevskian geometry.

We conclude this article with one remark. In assigning the roles of lines to equidistants, obtained by normal sections of the equidistant surface Σ , for geometry on this surface it was not proved that these equidistants are its geodesics. Now this can be easily done. Indeed, since all the theorems of absolute geometry are true for an equidistant surface, it may be shown in the usual manner that the segment

of an equidistant is shorter than any other curve joining their end-points on Σ .

§ 47. Now we will deal with geometry on a horosphere. The role of lines in this geometry is taken over by oricycles, which are obtained by intersecting the horosphere by a plane passing through an axis of the horosphere. (Here again the question as to whether these oricycles are geodesics of the horosphere is left open; an affirmative answer will be provided after investigating the geometry on the horosphere). Our present aim is to show that the mutual relationships between points and oricycles on the horosphere may be expressed with the aid of axioms of absolute geometry. Then there will be a discussion as to what the theory of parallels—Euclidean or Lobachevskian—is valid on a horosphere.

That Axioms I.1-I.3 of the first group are valid is disposed of easily. Let A and B be any two points on the horosphere and a and b two axes passing through these points. Since any two axes of the horosphere lie in one plane, the lines a and b determine just one plane α which contains them. The intersection of α with the given horosphere (denoted by Ω) determines just one oricycle u passing through A and B . Thus, given two points A and B on the horosphere Ω , they determine just one oricycle passing through them.

This shows that Axioms I.1-I.2 hold for geometry on the horosphere. It follows immediately from the definition of an oricycle and a horosphere and from simple theorems of Lobachevskian solid geometry that an oricycle has at least two points and a horosphere at least three points not lying on one oricycle (in fact, both contain an infinite number of points), that is, that Axiom I.3 is valid for geometry on a horosphere. (In order not to divert the attention of the reader with minor details, in discussing an oricycle and a horosphere no attention was paid to proving their obvious properties.)

Next it should be found whether Axioms II.1-II.4 (axioms of betweenness) are valid on a horosphere. First, we must clarify what is meant by a point of the horosphere lying between any other two of its points. Consider an oricycle u contained in a plane α and take points A , B , C on the oricycle. Point B will be said to lie between A and C if an axis of the oricycle, b , through B lies in α between axes a and c passing through A and C , respectively (that is, in α all the points of a and c lie on the opposite sides of b). It may be easily shown that linear axioms of betweenness (Axioms II.1-II.3) are fulfilled. What is somewhat more difficult to show is the validity of Pasch's axiom II.4. To do this, on the horosphere Ω consider any triangle ABC (Fig. 61) consisting of arcs of three oricycles. Through the vertices A , B , C draw three axes a , b , c , respectively, of Ω . Further, on these lines take the respective points A' , B' , C' and through them draw a plane σ . Pasch's axiom II.4 is valid for any triangle, hence also for triangle $A'B'C'$ in the plane σ . This will imply its validity

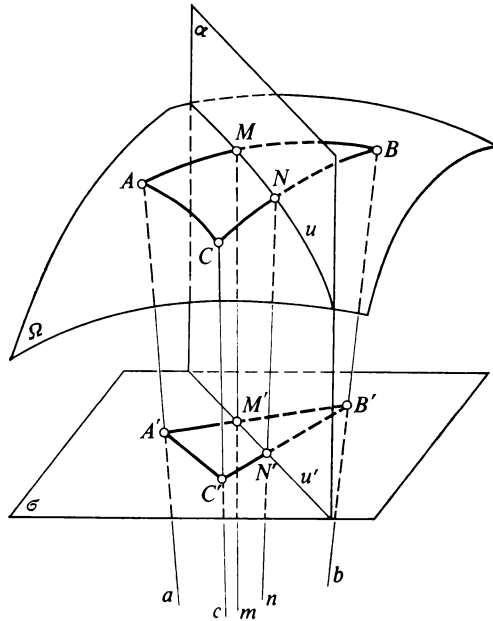


Figure 61

for triangle ABC on Ω also. Let u be some oricycle of Ω not passing through A, B, C . We have to show that if u passes through an interior point M of a segment of the oricycle AB , then u passes through either one of the interior points of the segment of the oricycle BC or one of the interior points of the segment AC . Note that the plane α , which contains u , and the plane $ABB'A'$ intersect along the axis m passing through M of the horosphere Ω . In the plane $ABB'A'$, this axis m lies between axes a and b because M lies between A and B . Therefore m must intersect rectilinear segment $A'B'$ at some interior point M' . Then, by Pasch's axiom, line u' along which planes α and σ intersect passes through the interior point of either the segment $B'C'$ or the segment $A'C'$. To be definite, suppose that u' contains interior point N' of segment $B'C'$. Then planes α and $BCC'B'$, having a common point N' , must intersect along some line, n . But planes α and $BCC'B'$ pass through two parallel lines m and b ; therefore, by Lemma IV, § 35, we find that line n , along which these two planes intersect, is parallel to each line m and b , and is hence also parallel to c . Since N' lies between B' and C' , the line n , being parallel to b and c , lies between them. On the other hand, like any

other line being parallel to the axes of oricycle BC and lying in its plane, n is also an axis of this oricycle and therefore must intersect it at some point N (see § 38). Since n lies between lines b and c , the point N on the oricycle BC lies between B and C .

As n lies in plane α , this plane also contains point N . Thus, among the points of intersection of plane α with horosphere Ω , that is, among the points of the oricycle u , there is an interior point of a segment of the oricycle BC . This proves the validity of Pasch's axiom for the geometry on a horosphere.

Let us now turn to Axioms III.1-III.5 (axioms of congruence).

Axiom III.1 requires that on any oricycle of a horosphere Ω from any point in a given direction a segment may be uniquely marked off which is congruent to any given segment of another oricycle. Axiom III.4 requires that on a horosphere Ω on any side of the given oricycle it is possible to join to the oricycle an angle which is congruent to any given angle; what is more, the vertex can be arbitrarily chosen and then the construction must be unique. Both these axioms are fulfilled in geometry on the horosphere in view of the fact that a horosphere admits motions about itself the collection of which is transitive relative to linear elements. The uniqueness of the required construction follows from Theorem B of § 19.

Next, Axiom III.2 is fulfilled thanks to the group property of motions (see § 19).

To show the validity of Axiom III.3 in geometry on a horosphere Ω , consider two oricycles u and u' on this surface. Take three points A, B, C on u with B between A and C , and let A', B', C' be three points on u' having an analogous position. If $AB \equiv A'B'$, there is a motion of the horosphere about itself which makes point A' coincide with point A and B' with B . If, moreover, $BC \equiv B'C'$, then Theorem B, § 19, implies that under this motion C' coincides with C . Thus under the motion in question segment $A'C'$ is superposed on segment AC , that is, if $AB \equiv A'B'$ and $BC \equiv B'C'$, then $AC \equiv A'C'$.

The proof of the validity of Axiom III.5 is also simple.

It remains to be seen whether Axioms IV.1-IV.2 (axioms of continuity) are also valid. Instead of showing separately the validity of Archimedes' axiom IV.1 and Cantor's axiom IV.2 for the geometry on a horosphere, it is more profitable to establish the validity of Dedekind's principle. If we do so, then this principle together with the axioms of Groups I-III will imply, in view of Theorem 41 of § 23, the validity of Axioms IV.1 and IV.2 for a horosphere.

On a horosphere, consider an oricycle u and let α be its plane. Let a Dedekind's cut be performed over the points of this oricycle. In the first class of this cut take any point A and in the second class a point B ; through these points draw axes a and b of the oricycle. Choosing a point A' on the first of these lines and a point B on the

second, we draw a line u' defined by A' and B' . Now note that through each point M' of line u' , just as through each point of plane α , there passes just one axis of oricycle u intersecting it at a point M . Thus to each point M' on u' there corresponds, according to the construction, a definite point M of the oricycle u . We divide all the points of u' into two classes such that point M' belongs to the first class if its corresponding point M of the oricycle u belongs to the first class of Dedekind's cut and to the second class if the corresponding point of the oricycle belongs to the second class. Obviously, this division of points of u' is a Dedekind's cut. Since Dedekind's principle holds for lines in Lobachevskian space, it can be asserted that one of the classes of Dedekind's cut performed over u' contains a closing element.

Let us denote this element by X' and assume (for the sake of definiteness) that this is the first point of the second class. Since A' and B' belong to different classes, X' lies between them or, in the extreme case, coincides with B' . The point X on the oricycle corresponding to X' belongs to the second class of Dedekind's cut performed over the points of the oricycle and lies between A and B or, in the extreme case, coincides with B . If X is not a closing element of the second class on the oricycle, then there is a point Y between A and X also belonging to the second class. Axis y of oricycle u , passing through Y , is situated between axes AA' and XX' ; hence it intersects segment $A'X'$ at some point Y' . Point Y' of u' belongs to the second class, since Y belongs to the second class on oricycle u . But at the same time the point Y' , by its construction, lies on u' closer to the points of the first class than X' . This is impossible, because X' is the first point of the second class; the contradiction shows that X necessarily closes the second class. Had it been assumed that X' on u' closes the first class, it could be proved similarly that its corresponding point X closes the first class of Dedekind's cut on the oricycle.

Thus, given a Dedekind's cut on an oricycle, one of the classes of this cut has necessarily a closing element. This proves that Dedekind's principle holds in geometry on the horosphere. Theorem 41 of § 23 then implies that axioms of continuity holds for a horosphere.

The above analysis enables us to conclude that all the propositions of absolute geometry are valid for a horosphere. Indeed, all of them may be obtained in a logical way from the axioms of Groups I-IV, which have been shown to be valid.

Now we must answer the question: Which theory of parallels—Euclidean or Lobachevskian—is realized in the geometrical system on a horosphere? An answer to this question can be easily provided.

On horosphere Ω take any oricycle a and denote its plane by α . Let P be any point on the horosphere Ω outside a . Draw through P an axis p of the horosphere. Theorem XII of § 35 states that p

is parallel to α . Suppose that an arbitrary oricycle b is drawn through P . Its plane β goes through p . Oricycle b will have no common point with oricycle a only if β and α do not intersect. But since p is parallel to α , it follows from Theorem XIV of § 35 that just one plane β passes through p not intersecting α . Therefore through P there is just one oricycle on horosphere Ω not intersecting oricycle a . Thus on a horosphere Euclid's parallel postulate holds.

Thus one may assert that elementary geometry on a horosphere is Euclidean geometry. This remarkable result plays an important role in the development of Lobachevskian geometry. This fact, irrespective of its use, is of interest in its own right. It turns out that if Euclid's fifth postulate is rejected in the two-dimensional geometry on every plane, we yet recover this postulate in the two-dimensional geometry on some other surface.

It is interesting to compare geometries on an equidistant surface, on a horosphere and on an ordinary sphere by considering in these geometries the proposition regarding the angle-sums of triangles.

Since geometry on an equidistant surface is Lobachevskian geometry, it follows that the angle-sum of any triangle formed by the arcs of the geodesics (that is, by arcs of equidistants) is less than two right angles.

Since Euclidean geometry is realized on a horosphere, the angle-sum of the geodesic triangle (formed by arcs of oricycles) on it is equal to two right angles.

The spherical triangle whose sides are arcs of great circles (that is, of geodesics of a sphere) has an angle-sum greater than two right angles. On a sphere, there is a geodesic triangle with three right angles.

Thus in spherical geometry the very proposition is valid whose falsity was established for absolute geometry by many geometers (Legendre, Saccheri, Lambert; the latter two in the form of the obtuse angle hypothesis). The reason for this is of course that spherical geometry is far less similar to Euclidean plane geometry than to Lobachevskian plane geometry.

In fact, for spherical geometry not only is the Euclidean parallel axiom not valid, but a majority of axioms of absolute geometry is also not valid (for example, two geodesics of a sphere always meet in two points, the "betweenness" notion cannot be applied to points on geodesics, etc.).

We conclude with the remark that in one respect Lobachevskian space is richer than Euclidean. Namely, whereas in Euclidean space there are only two elementary geometries of two-dimensional manifolds—spherical and Euclidean—in Lobachevskian space all three geometric systems known to us are realized on various surfaces.

3.8. Area of a triangle

§ 48. In the preceding section, we examined only angular and linear magnitudes. Now the problem of determining the area of figures in the Lobachevskian plane will be dealt with.

In defining area, the notion of the equicomposition of figures is used: two figures are said to be *equicomposed* if they can be divided into the same number of mutually congruent triangles. For some time we confine our discussion to triangles only.

The following proposition holds: a necessary and sufficient condition for the equicomposition of two triangles is that their defects coincide. We recall that the defect of a triangle $\triangle$ is the difference

$$D(\triangle) = \pi - S(\triangle)$$

where $S(\triangle)$ is the angle-sum of the triangle; according to Legendre's theorem (Proposition 3 of § 8), $S(\triangle) < \pi$ and $D(\triangle) > 0$ in non-Euclidean geometry.

The proof of the necessity for the abovementioned criterion is based on the following two lemmas.

Lemma I. *Suppose that a simple connected region bounded by a closed broken line is divided into triangles, so that the following conditions hold: any two triangles in the division have either no common points or have one common vertex or one common side. Then, if α^2 denotes the number of triangles in the division, α_i^0 the number of vertices of the triangles lying inside the region and α_e^0 the number of vertices on the boundary,*

$$\alpha^2 - 2\alpha_i^0 - \alpha_e^0 = -2 \quad (\text{A})$$

(In Fig. 62, $\alpha^2 = 10$, $\alpha_i^0 = 3$, $\alpha_e^0 = 6$.)

Proof. We use the well-known Euler's formula

$$\alpha^2 - \alpha^1 + \alpha^0 = 1$$

where α^1 is the total number of the sides of the triangles in the division and α^0 the total number of vertices.* Let the vertices of the triangles in the division be numbered in some way, and let p_{ik}^2 denote the number of triangles having a common interior vertex with the triangle number k and p_{er}^2 the number of triangles with a common vertex on the boundary with the triangle number r . Denote by p_{ik}^1 and p_{er}^1 the number of sides originating from these vertices. Then, obviously

$$\left. \begin{aligned} p_{ik}^2 &= p_{ik}^1 \\ p_{er}^2 &= p_{er}^1 - 1 \end{aligned} \right\} \quad (\text{B})$$

* See, for instance, W. G. Chinn and N. E. Steenrod: *First Concepts of Topology*, Random House, New York, 1966.

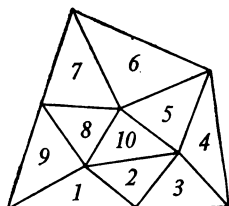


Figure 62

On the other hand, summing over all the interior and exterior vertices, we have

$$\sum_k p_{ik}^2 + \sum_r p_{er}^2 = 3\alpha^2$$

$$\sum_k p_{ik}^1 + \sum_r p_{er}^1 = 2\alpha^1$$

Subtracting the lower equation from the upper and noting (B), we obtain

$$\alpha_e^0 = 2\alpha^1 - 3\alpha^2$$

Eliminating α^1 from this equation and from Euler's formula

$$\alpha^2 - \alpha^1 + \alpha^0 = 1$$

we obtain

$$\alpha^2 - 2\alpha^0 + \alpha_e^0 = -2$$

Noting that $\alpha^0 = \alpha_i^0 + \alpha_e^0$ and substituting it for α^0 in the last equation, we have the desired result

$$\alpha^2 - 2\alpha_i^0 - \alpha_e^0 = -2 \quad \blacksquare$$

In topology, the division of regions by triangles satisfying the conditions of Lemma I is called *triangulation*.

Lemma II. *If a triangle $\triangle$ is composed of triangles $\triangle_1, \triangle_2, \dots, \triangle_n$, then*

$$D(\triangle) = D(\triangle_1) + \dots + D(\triangle_n)$$

Proof. This lemma is evidently a generalization of Lemma 3 of § 8, according to which if a transversal BD divides a triangle ABC into two triangles ABD and BDC , then

$$D(ABC) = D(ABD) + D(BDC)$$

At the same time, Lemma 3 of § 8 implies that the proof of Lemma II can be reduced to the case where the division of triangle $\triangle$ constitutes a triangulation.

Indeed, the triangles $\triangle_1, \triangle_2, \dots, \triangle_n$ do not satisfy the triangulation conditions if the vertices of some triangles ∇_i coincide with the inte-

rior points of the sides of some other triangles $\triangle_i$. But by joining successively the vertices of triangles $\triangle_i$ that lie on the sides of neighbouring triangles with the vertices opposite to these sides of the latter triangles $\triangle_i$, we obtain a new system of triangles: $\triangle'_1, \dots, \triangle'_m$. The division of triangle $\triangle$ into triangles $\triangle'_1, \dots, \triangle'_m$ is a triangulation. Furthermore, the sum of the defects of $\triangle'_1, \dots, \triangle'_m$ is equal to the sum of the defects of $\triangle_1, \dots, \triangle_n$, since each time a transversal divides triangle $\triangle_i$ this results in two triangles the sum of whose defects, according to Lemma 3 of § 8 equals the defect of $\triangle_i$.

Thus for the proof of the lemma, it suffices to show that

$$D(\triangle'_1) + \dots + D(\triangle'_m) = D(\triangle)$$

If l denotes the number of vertices of triangles $\triangle'_1, \dots, \triangle'_m$ lying inside $\triangle$, and p denotes the number of vertices lying on the sides of triangle $\triangle$ (the three vertices of $\triangle$ are not counted), then

$$m - 2l - p = 1 \quad (C)$$

This equation is obtained by a light modification of formula (A) in the preceding lemma. In fact, applying Lemma I to the division of $\triangle$ into triangles $\triangle'_1, \dots, \triangle'_m$, we have

$$\alpha^2 = m, \quad \alpha_i^2 = l, \quad \alpha_e^2 = p + 3$$

Substitution of this in (A) results in (C).

Consider now the sum $D(\triangle'_1) + \dots + D(\triangle'_m)$. Obviously,

$$D(\triangle'_1) + \dots + D(\triangle'_m) = m\pi - [S(\triangle'_1) + \dots + S(\triangle'_m)]$$

The angle-sum of the triangles $\triangle'_1, \dots, \triangle'_m$ meeting at every common vertex lying inside $\triangle$ is 2π ; the sum of angles bordering every vertex lying on the side of $\triangle$ is π , and, finally, the angle-sum of triangles $\triangle'_1, \dots, \triangle'_m$ at the vertices which coincide with the vertices of $\triangle$ is $S(\triangle)$. Hence

$$S(\triangle'_1) + \dots + S(\triangle'_m) = 2l\pi + p\pi + S(\triangle)$$

This yields

$$D(\triangle'_1) + \dots + D(\triangle'_m) = (m - 2l - p)\pi - S(\triangle)$$

or, noting (C),

$$D(\triangle'_1) + \dots + D(\triangle'_m) = \pi - S(\triangle) = D(\triangle)$$

But since

$$D(\triangle'_1) + \dots + D(\triangle'_m) = D(\triangle_1) + \dots + D(\triangle_n)$$

it follows that

$$D(\triangle_1) + \dots + D(\triangle_n) = D(\triangle) \quad \blacksquare$$

The following theorem proves the necessity for the above equicomposition criterion for triangles.

Theorem I. *Equicomposed triangles have equal defects.*

Proof. Assume that the triangles $\triangle$ and $\triangle'$ have been divided into the same number of pairwise congruent triangles $\triangle_1, \dots, \triangle_n$ and $\triangle'_1, \dots, \triangle'_n$; the triangles are numbered so that triangles $\triangle_i$ and $\triangle'_i$ with the same numbers are congruent. By Lemma II,

$$D(\triangle) = D(\triangle_1) + \dots + D(\triangle_n)$$

and

$$D(\triangle') = D(\triangle'_1) + \dots + D(\triangle'_n)$$

But since congruent triangles have evidently equal defects, it follows that

$$D(\triangle_i) = D(\triangle'_i)$$

which together with the preceding equations implies that

$$D(\triangle) = D(\triangle') \quad \blacksquare$$

The sufficiency for the equicomposition criterion for triangles is given in

Theorem II. *If two triangles have the same defects, they are equicomposed.*

This theorem follows from a number of lemmas proved below.*

Lemma α . *Two figures equicomposed with a third are equicomposed among themselves.*

Proof. Suppose that figures A and B are equicomposed with a third figure C . Suppose also that on both figures A and B those lines are marked which divide these figures into parts congruent to parts of C . Plot lines on C which divide it into parts that are respectively congruent to parts of A and then plot lines dividing C into parts that are respectively congruent to parts of B . Then all these lines divide C into parts from which both A and B can be formed. $\blacksquare$

Lemma β . *If E and F are the feet of the perpendiculars drawn from vertices B and C of a triangle ABC to the line joining the midpoints P and Q of the sides AB and AC , then $BCFE$ is a Saccheri quadrilateral and triangle ABC is equicomposed with the quadrilateral $BCFE$.*

Proof. We must first show that $BCFE$ is a Saccheri quadrilateral. Draw from A a perpendicular AD to line PQ ; obviously, $\triangle BEP = \triangle ADP$ and $\triangle CFQ = \triangle ADQ$, whence $BE = AD$ and $CF = AD$. Accordingly, $BE = CF$ and thus $BCFE$ is in fact a Saccheri quadrilateral. To prove that $\triangle ABC$ and quadrilateral $BCFE$ are equicomposed, we discuss two cases.

(1) Segment PQ is a part of segment EF (Fig. 63a and b).

That ABC and $BCFE$ are equicomposed in this case follows immediately from Fig. 63a and b, where equal triangles have been indi-

* These lemmas have been partly taken from R. Baldus: *Nichteuklidische Geometrie. Hyperbolische Geometrie der Ebene*, De Gruyter, Berlin-Leipzig, 1927.

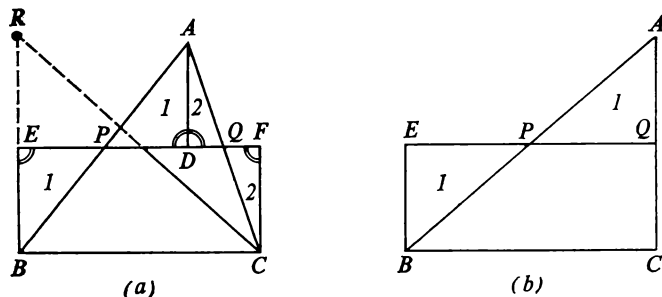


Figure 63

cated by identical numbers (Fig. 63b refers to the case where F and Q coincide).

(2) The segment PQ lies, at least partly, outside EF (Fig. 64). In this case note first that $PQ = (1/2) EF$. This follows from the fact that

$$\triangle BEP = \triangle ADP \quad \text{and} \quad \triangle CFQ = \triangle ADQ$$

implying that $EP = PD$ and $FQ = QD$, which in turn yields $EP - FQ = PD - QD$ or $EF - PQ = PQ$, that is, $2PQ = EF$ or $PQ = (1/2) EF$.

Next join C with point P and on this line mark off segment $PA' = PC$. Then join A' with B , and let P' be the point of intersection of line EF with side BA' of $\triangle A'BC$. It is easy to show that P' is the midpoint of side $A'B$. Indeed, if $A'D'$ is the perpendicular from A' to EF , then $A'D' = CF = BE$; therefore $\triangle P'A'D' = \triangle P'BE$, whence $BP' = P'A'$. Furthermore, triangles $A'BC$ and ABC are equicomposed since they contain a common triangle BPC , while triangles BPA' and CPA are equal because the angles included between the respective equal sides are equal. Thus, starting with the triangle ABC , one can construct a triangle $A'BC$ in the manner described; similarly, starting with the triangle $A'BC$, one can form a new triangle $A''BC$, and so on (Fig. 64).

All the triangles ABC , $A'BC$, $A''BC$, ... have a common median and are, in view of the previous result and Lemma α , equicomposed. Moreover,

$$QP = PP' = P'P'' = \dots = (1/2) EF$$

By virtue of Archimedes' axiom, one of these triangles is situated as in the first part of the proof; accordingly, it is equicomposed with the Saccheri quadrilateral $BCFE$ and hence the quadrilateral $BCFE$ and the original triangle ABC are equicomposed. ■

Lemma γ . *If two triangles have the same defects and one side of one of the triangles is equal to one side of the second triangle, then the Saccheri quadrilaterals corresponding to these sides are congruent.*

Proof. Indeed, $\angle B$ and $\angle C$ of the quadrilateral BCE are equal and, as is clear from Fig. 63 or Fig. 64, each of them is $S/2$, where S is the angle-sum of triangle ABC . But a Saccheri quadrilateral is completely determined by the top side and the summit angle. ■

Lemmas α , β and γ immediately imply

Lemma δ . *If two triangles have the same defects and one side of one of the triangles is equal to one side of the other, then the two triangles are equicomposed.*

If segment BE in Fig. 63 is produced to a point R such that $BE = ER$, then evidently the midpoint of side RC of triangle BRC will lie on the line EF . According to Lemma β , the triangles BRC and BAC are equicomposed with the quadrilateral $BCEF$, and hence, by Lemma α , they are equicomposed. Moreover, $\angle EBC = S/2$. Thus we have proved

Lemma ε . *If the angle-sum of a triangle ABC is S , then an equicomposed triangle can be constructed with one of its angles being $S/2$.*

Consider now two triangles with the same defects, and hence with the same angle-sums S . According to Lemma ε , one can construct two new triangles equicomposed with respective triangles, and in each of the new triangles one angle is $S/2$. We superpose these triangles so that equal angles coincide (Fig. 65). If the vertices C and C_1 coincide, then vertices A and A_1 also must coincide, since otherwise one triangle will become a part of the other and, by Lemma 3 of § 8 that triangle will have a smaller defect, contrary to our assumption. In this case all the vertices of the triangles will coincide and the triangles will be congruent. If all the vertices do not coincide, then, for the same reason, one triangle cannot lie wholly inside the other, and the triangles will be situated as shown in Fig. 65.

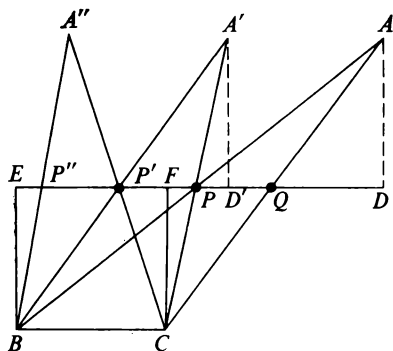


Figure 64

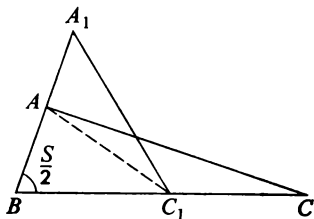


Figure 65

Drawing line AC_1 , we find from Lemma 3 of § 8 that the triangles AC_1A_1 and AC_1C have the same defects and hence, by Lemma δ , they are equicomposed. Therefore triangles ABC and A_1BC_1 are also equicomposed. This completes the proof of Theorem II. ■

Now consider triangle ABC . Taking a point D on the side BC , set $BC = a$ and $BD = x$. The defect of triangle BAD is evidently a function of x :

$$D(ABD) = D(x)$$

If we agree to take $D(0) = 0$, the function $D(x)$ is defined for all values of x , $0 \leq x \leq a$. We will now show that $D(x)$ is continuous for all x ($0 \leq x \leq a$).

Denote $\angle BAD$ by $\alpha(x)$ and $\angle BDA$ by $\beta(x)$. It suffices to establish the continuity of $\alpha(x)$ and $\beta(x)$. Since these functions are monotone, their continuity will be proved if it is shown that besides assuming any two given values they also assume the intermediate values.

For the function $\alpha(x)$ this is obvious, as the line AD can be drawn at any angle of inclination to AB . It is also easy to see that the angle of inclination of AD to BC assumes all possible values (between 0 and π if the position of D is unrestricted on BC). Indeed, consider $\triangle MON$ with an arbitrary magnitude β_0 and from a variable point M^* draw perpendicular M^*N^* to ON on the side OM of the angle. Set $OM^* = x^*$ and $M^*N^* = y^*$. By Lemma II of § 30,

$$y^* = f(x^*)$$

is a continuous function that increases monotonically without bounds. This implies that there is a value y_0^* equal to the length of the altitude of triangle ABC drawn on the side BC . Denote the corresponding position of M^* and N^* by M_0^* and N_0^* . We now place triangle $OM_0^*N_0^*$ in such a way that M_0^* coincides with A and side $M_0^*N_0^*$ lies on the altitude drawn from A to side BC of the triangle ABC . The points O and N_0^* will evidently lie on line BC . Denote by D_0 the point from which O moves and by x_0 the length of BD_0 . By its construction, $\beta(x_0)$ has the preassigned value β_0 . Thus our assertion is proved; namely, $D(x)$ is continuous for $0 \leq x \leq a$.

Having established the continuity of $D(x)$, we can state the following

Theorem III. *Given any number α satisfying the inequalities $0 < \alpha \leq n^{-1}D(ABC)$, there is a system of points $D_1, D_2, \dots, D_n$ on side BC of the triangle ABC such that each of the triangles $BAD_1, D_1AD_2, \dots, D_{n-1}AD_n$ has a defect α .*

The proof follows from the continuity of the defect.

Now we are in the position to define the area of a triangle.

In Euclidean geometry, the area of a triangle is defined in such a way that the following two conditions hold: (i) congruent triangles have the same area, and (ii) if a triangle $\triangle$ is composed of triangles $\triangle_1, \dots, \triangle_n$, then the area of $\triangle$ is equal to the sum of the areas of $\triangle_1, \dots, \triangle_n$.

These two conditions will serve as the basis for defining the area of a triangle in non-Euclidean geometry.

Suppose that with each triangle $\triangle$ in the Lobachevskian plane there is associated a positive number $f(\triangle)$; in other words, a function $f(\triangle)$ is defined whose domain of definition is the set of all triangles and all its values are positive. Furthermore, assume that the following conditions are satisfied:

- (i) if triangles $\triangle_1$ and $\triangle_2$ are equal, then $f(\triangle_1) = f(\triangle_2)$;
- (ii) if triangle $\triangle$ is composed of triangles $\triangle_1, \dots, \triangle_n$, then

$$f(\triangle) = f(\triangle_1) + \dots + f(\triangle_n)$$

Then the number $f(\triangle)$ will be called the *area of triangle $\triangle$* .

For this definition to make sense, the existence of a function $f(\triangle)$ satisfying conditions (i) and (ii) must be established. We will show that such a function $f(\triangle)$ exists and is "unique" in the sense that it is completely defined by prescribing its value for a given triangle; in other words, if one triangle is assigned a definite area, then the area of every triangle is completely determined.

As to the question of existence, it has been solved in the previous discussion: the defect $D(\triangle)$ of the triangle $\triangle$ has properties (i) and (ii). The question concerning the uniqueness of the value of area is settled by the following theorem.

Theorem IV. *Any function $f(\triangle)$ satisfying conditions (i) and (ii) has the form*

$$f(\triangle) = kD(\triangle) \quad (*)$$

where k is a positive constant, that is, is a positive number independent of $\triangle$.

Indeed, if this theorem is assumed to be valid, then, fixing a value of function $f(\triangle)$ for some triangle $\triangle_0$, we completely determine this function because the relation

$$f(\triangle_0) = kD(\triangle_0)$$

completely determines the value of k .

It may be said that the choice of one of the functions $f(\Delta)$ as the area of a triangle or, what is the same, the choice of constant k in (*) corresponds to assigning a definite number to areas. It should be noted, however, that if the function $f(\Delta)$ is chosen as the area arbitrarily, there may not be any triangle with area equal to unity. For example, if $k < 1/\pi$, then for all triangles $f(\Delta) < 1$ since the defect of every triangle is less than π .

We now prove Theorem IV. It suffices to show that for any two triangles Δ and $\bar{\Delta}$

$$\frac{f(\Delta)}{f(\bar{\Delta})} = \frac{D(\Delta)}{D(\bar{\Delta})}$$

because by fixing triangle $\bar{\Delta}$ and setting $f(\bar{\Delta})/D(\bar{\Delta}) = k$ we obtain (*).

Take any positive integer n and divide triangle $\bar{\Delta}$ into triangles $\bar{\Delta}_1, \bar{\Delta}_2, \dots, \bar{\Delta}_n$ by transversals originating from some vertex of triangle $\bar{\Delta}$ such that their defects are the same; then

$$D(\bar{\Delta}_i) = \frac{D(\bar{\Delta})}{n} \quad (i = 1, 2, \dots, n)$$

Denoting the vertices of Δ by A, B, C , take a system of points $A_1, A_2, \dots, A_m$ on the side AC subjecting them to the following condition: if Δ_1 denotes the triangle ABA_1 , Δ_2 the triangle A_1BA_2 , and so on, then (a) the defects of all the triangles $\Delta_1, \Delta_2, \dots, \Delta_m$ are $D(\Delta)/n$, (b) either point A_m coincides with point C or

$$D(A_mBC) < \frac{D(\bar{\Delta})}{n}$$

That such a system can be chosen is ensured by Theorem III.

Indeed, if m denotes the largest natural number satisfying

$$mD(\bar{\Delta}) \leq nD(\Delta)$$

then

$$\alpha \leq \frac{D(\Delta)}{m}$$

if we put

$$\alpha = \frac{D(\bar{\Delta})}{n}$$

According to Theorem III, points $A_1, \dots, A_m$ will be found on the side AC of triangle ABC such that they correspond to triangles $\Delta_1, \dots, \Delta_m$ with defects α . Moreover, A_m coincides with C if $\alpha = D(\Delta)/m$ and precedes C if $\alpha < D(\Delta)/m$; in the latter case the defect of triangle A_mBC would be less than α , otherwise we would have

$(m+1) D(\bar{\Delta}) \leq nD(\Delta)$, violating the condition. Obviously

$$\frac{m}{n} D(\bar{\Delta}) \leq D(\Delta) < \frac{m+1}{n} D(\bar{\Delta})$$

or

$$\frac{m}{n} \leq \frac{D(\Delta)}{D(\bar{\Delta})} < \frac{m+1}{n}$$

whence

$$\lim_{n \rightarrow \infty} \frac{m}{n} = \frac{D(\Delta)}{D(\bar{\Delta})} \quad (\text{E})$$

Note next that since the triangles $\bar{\Delta}_1, \bar{\Delta}_2, \dots, \bar{\Delta}_n, \Delta_1, \Delta_2, \dots, \Delta_m$ have the same defects, Theorem II implies that all of them are equicomposed with one of them. This together with conditions (a) and (b) implies that

$$f(\bar{\Delta}_1) = \dots = f(\bar{\Delta}_n) = f(\Delta_1) = \dots = f(\Delta_m)$$

or

$$\begin{aligned} f(\bar{\Delta}_i) &= \frac{f(\bar{\Delta})}{n} & (i=1, 2, \dots, n) \\ f(\Delta_j) &= \frac{f(\bar{\Delta})}{n} & (j=1, 2, \dots, m) \end{aligned} \quad (**)$$

Further, by condition (b),

$$mf(\Delta_j) \leq f(\Delta) < (m+1)f(\Delta_j)$$

This, on the basis of (**), yields

$$\frac{m}{n} f(\bar{\Delta}) \leq f(\Delta) < \frac{m+1}{n} f(\bar{\Delta})$$

or

$$\frac{m}{n} \leq \frac{f(\Delta)}{f(\bar{\Delta})} < \frac{m+1}{n}$$

Accordingly,

$$\lim_{n \rightarrow \infty} \frac{m}{n} = \frac{f(\Delta)}{f(\bar{\Delta})} \quad (\text{F})$$

Comparing (E) and (F), we have

$$\frac{f(\Delta)}{f(\bar{\Delta})} = \frac{D(\Delta)}{D(\bar{\Delta})} \quad \blacksquare$$

Thus we have shown that conditions (a) and (b) determine the area $f(\Delta)$ to within a constant multiple:

$$f(\Delta) = kD(\Delta) \quad (\text{I})$$

In § 182 a connection will be established between the choice of units measuring area and length (in Euclidean geometry, such a connection is provided by the fact that the unit of area is chosen as the area of a square whose sides are equal to the unit of length). In this way k will be fixed together with the choice of a linear scale.

Having defined the area of a triangle, defining the area of any polygon offers no complications: assuming any polygon P divided into triangles $\triangle_1, \dots, \triangle_n$, we define the area of P as a number σ equal to the sum of the areas of $\triangle_1, \dots, \triangle_n$. The reader can easily show that the value of σ does not depend on the manner in which a polygon is divided into triangles.

A few remarks concerning the above presentation are in order. Since, by its definition, the defect of a triangle is less than π , the area of each triangle is less than $k\pi$. Thus a theorem may be stated: in absolute geometry the assumption that there exists a triangle with an arbitrarily large area is equivalent to Euclid's fifth postulate. Indeed, as is clear from the above discussion, this assumption is not valid in the Lobachevskian system.

On the other hand, since there are polygons which cover an arbitrarily large number of similar triangles, the areas of polygons can be as large as possible. What is more, the continuity of the defect implies that there is a polygon whose area is any preassigned positive number. In particular, there is a polygon with unit area.

In conclusion, we compare the area in Lobachevskian geometry with the area in spherical geometry. It is known that the area of a spherical triangle is given by

$$\sigma(\triangle) = R^2(\alpha + \beta + \gamma - \pi) \quad (\text{II})$$

where R is the radius of the sphere and α, β, γ the angles of the triangle. But formula (I) may also be expressed as

$$\sigma(\triangle) = k(\pi - \alpha - \beta - \gamma) \quad (\text{I}')$$

This shows that (I') is obtained from (II) if the radius R of the sphere is replaced by an imaginary quantity $i\sqrt{k}$. This feature was noted by Lambert.

3.9. Proof of logical consistency of Lobachevskian geometry

§ 49. We have established the basic facts of the parallel theory in Lobachevskian geometry. Although many of these facts contradict the properties of lines as we know them, we were not able to find any logical inconsistency in the preceding exposition. On the other hand, non-Euclidean geometry, at least the portion known to us, has the form of a theory nicely constructed in the logical sense.

But what is the guarantee that in developing this theory further we shall not encounter logical inconsistencies? Lobachevsky clearly understood that in order to show the independence of Euclid's fifth postulate from other geometric postulates it is not enough to produce a set of theorems obtained by denying the validity of the Euclidean postulate and then declare that there is no logical inconsistency in the new set. He felt that some sort of argument was required which would show that the accepted statements would never lead to contradictions, that is, that the Euclidean postulate can never be proved by the method of contradiction.

By obtaining the basic equations of his geometry, Lobachevsky gave an analytical interpretation of this geometry and thereby showed, in principle, its consistency. Later (towards the end of the nineteenth century), when the understanding of geometrical objects and geometric axioms was widened, the consistency of Lobachevskian geometry was shown in a simple and precise manner; to one of these proofs, due to the French mathematician Jules Henri Poincaré (1854-1912), the next few pages will be devoted.

So as not to obscure the exposition by difficulties of a technical nature, we confine the discussion to two-dimensional geometry. The problem then posed can be formulated as follows: show that Axioms I.1-I.3, the axioms of Groups II-IV and the non-Euclidean parallel axiom are logically consistent, that is, these axioms do not lead to any two assertions negating each other.

The general idea of the solution is indicated by the modern view of geometric axioms. Going back to § 11, where geometric objects were introduced, we find that there is not even a hint of a description of geometric elements: points, lines and planes; only the existence of certain objects called by these names is assumed. Further, between these elements there are certain relationships expressed by "lies on ...", "between", "are congruent"; but the nature of these relationships is not given, they are only assumed to have a certain number of properties (very small number, to that) mentioned in axioms.

Therefore in studying, for example, Euclidean plane geometry, by "point" and "line" we mean any specific objects and express by "lies on ...", "between", "are congruent" any relationships among these objects, the only condition being that they be consistent with Axioms I.1-I.3 and the axioms of Groups II-V. Every proposition derived logically from Axioms I.1-I.3 and the axioms of Groups II-V expresses a certain fact relating to the chosen object. Clearly, the specific meaning of every abstract geometric proposition depends on the system of objects chosen. By choosing specific objects the relationships between which satisfy the requirements of the given system of axioms, we obtain a *model* of the abstract scheme defined by these axioms.

In the preceding section, where elementary geometry on an equidistant surface was examined, we had examples of various models of the same abstract scheme in Lobachevskian plane geometry. Indeed, as we know, both the relationships between points and equidistants on any equidistant surface and the relationships between points and lines in every plane of Lobachevskian space satisfy axioms of non-Euclidean plane geometry. However, we do not yet know whether such objects exist, since the very topic of our present discussion is the question of the existence of Lobachevskian space.

In constructing a concrete model of the Lobachevskian logical scheme lies the proof of its logical consistency.

The principle behind such a proof can be easily explained by examining a problem opposite the problem at hand. Suppose by some means we are able to show the consistency of Lobachevskian geometry and then pose the question of the consistency of Euclidean planimetry. This question could have been easily settled by examining a horosphere. On a horosphere the points and oricycles have the same sort of relationships as required by the axioms of Euclidean plane geometry (see § 47). Therefore if the axioms of Euclidean plane geometry lead to two contradictory results, this would imply the inconsistency of elementary geometry on a horosphere, that is, of Lobachevskian geometry because a horosphere is an object of this geometry.

Because a model of Euclidean plane geometry may be constructed in Lobachevskian space, the consistency of Lobachevskian geometry implies the consistency of Euclidean plane geometry.

Our aim is to prove the consistency of Lobachevskian geometry. In solving this problem, the consistency of Euclidean geometry is assumed (this question will be dealt with in the next chapter). Although there are no surfaces in Euclidean space whose elementary geometry coincides with Lobachevskian plane geometry, we can still construct a model of non-Euclidean planimetry from the objects of Euclidean space. But by speaking of points and lines, we shall only be forced farther from the obvious notions aroused by these words than if we consider elementary geometry of certain surfaces.

What is more, a model of non-Euclidean geometry can be constructed on the Euclidean plane and thus the consistency of Euclidean plane geometry would imply the consistency of Lobachevskian plane geometry.

The result which we shall obtain may be formulated thus: if the system of axioms of Euclidean plane geometry I.1-I.3, II-V is consistent, so is the system consisting of Axioms I.1-I.3, the axioms of Groups II-IV and the Lobachevskian parallel axiom. At the same time it will be shown that the Euclidean parallel axiom is not a necessary consequence of Axioms I.1-I.3 and the axioms of Groups II-

IV. We construct below a model or, as it is sometimes also called, an interpretation of non-Euclidean planimetry on the Euclidean plane due to Poincare.

§ 50. In the Euclidean plane consider a line x , assuming it to be horizontal for the sake of convenience. Line x determines two half planes, one of which will be called the upper. Points of the upper half plane (excluding the points of x) will be called *non-Euclidean points*, and by *non-Euclidean lines* we shall mean Euclidean semicircles lying in the upper half plane and orthogonal to x (that is, having centres on x) as well as Euclidean half lines in the upper half plane supported by line x and forming with it a right angle. In order to simplify the statements, these half lines will be referred to as semicircles of infinitely large radius.

Between non-Euclidean points and lines we establish relationships satisfying Axioms I.1-I.3 and the axioms of Groups II-IV, that is, satisfying axioms of absolute geometry. After that we will show that the Lobachevskian parallel axiom is realized in the system of objects thus constructed. The relationships between the objects will be established one after another as required by successive consideration of the axioms.

First we examine the axioms of Group I. To this group there precedes an assumption stating that geometric objects are in correspondence to each other as expressed by saying "a point lies on a line", "a line passes through a point", and so on. We must clarify what is to be understood by these expressions in relation to non-Euclidean points and lines.

Let A be a non-Euclidean point, and a a non-Euclidean line represented by a certain semicircle (the latter is also denoted by a). We will say that A lies on (non-Euclidean) line a if this point is on Euclidean semicircle a in the sense of the relationships of Euclidean geometry.

That Axioms I.1-I.3 are valid for non-Euclidean points and lines can be shown by means of Euclidean geometry.

Axiom I.1 is satisfied because through two points A and B of the upper half plane there always passes a semicircle orthogonal to line x .

Axiom I.2 is valid since two semicircles representing non-Euclidean lines can have at most one common point.

Axiom I.3 is true because there are infinitely many points on the semicircle and in the upper half plane there are infinitely many points not on one semicircle.

Consider now the axioms of Group II, the axioms of betweenness. We must determine what exactly is meant by the phrase "lies between..." in connection with the non-Euclidean points of a non-Euclidean line.

Let A , B , C be three point of a non-Euclidean line in the form of a semicircle a . Point B is said to lie (in the non-Euclidean sense)

between A and C if B lies on a between A and C in the sense of Euclidean geometry. In other words, the order of points on a non-Euclidean line coincides with the order of points on the Euclidean semicircle representing this line in the upper half plane.

A more detailed definition of the order of the points of a non-Euclidean line for the case where the semicircle representing this line does not degenerate into a Euclidean half line may be given in the following terms. Let a non-Euclidean line be represented by the semicircle a whose centre is at O (O is not an object of our system). Consider a Euclidean line u parallel to x . Each Euclidean line passing through O , except line x , intersects the semicircle a in one point M and the line u in one point M' , to be called a point corresponding to M .

Then, if A, B, C are three points of the non-Euclidean line represented by the semicircle a , the point B , being an object of non-Euclidean geometry, lies between A and C if in the system of points A', B', C' corresponding to points A, B, C on the Euclidean line u point B' lies between A' and C' .

This immediately implies that for a non-Euclidean line, Axioms II.1-II.3 are true, since these axioms are true for any Euclidean line.

We note in passing one important feature: Dedekind's principle is realized in the ordered set of points on a non-Euclidean line.

Indeed, Dedekind's principle is valid in Euclidean geometry. But as we have seen, between the points of a Euclidean line and the points of a non-Euclidean line a one-to-one correspondence may be set up such that the corresponding points are in the same order. And this proves the stated assertion.

Apart from Axioms II.1-II.3, whose validity has already been established, Group II also contains Pasch's axiom, Axiom II.4. To demonstrate that Pasch's axiom remains valid for our system, the following Euclidean theorem must be proved. Let ABC be a curvilinear triangle (Fig. 66) formed by the arcs of semicircles and a the semicircle not passing through A, B , or C , then if a passes through any point of the arc AC , it also passes either through a point of arc AB or through a point of arc BC . The proof of this theorem, being quite obvious, is of no interest.

The proofs of the validity of the axioms of the first two groups were reduced to establishing a number of trivial propositions of Eucli-

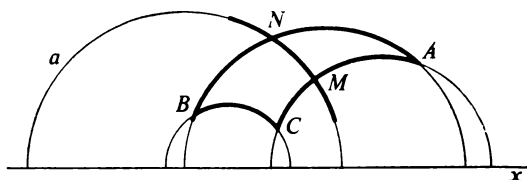


Figure 66

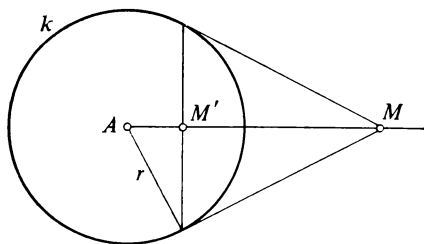


Figure 67

dean geometry. For more complicated is the verification of axioms of congruence III.1-III.5, the investigation of which will soon be started. This requires a proper definition of congruence of images.

§ 51. The main tool for our further construction will be a special mapping of the Euclidean plane onto itself, known as *inversion* in elementary geometry, in the theory of functions and in mathematical physics.

Consider a circle k with centre at A and radius r (Fig. 67). Let M be any point in the plane. If this point does not coincide with A , a unique point M' can always be found which is defined on the ray AM thus:

$$AM' \times AM = r^2 \quad (*)$$

(see the construction in Fig. 67). Point M' is called the *image* of M under inversion with respect to circle k or simply the *inverse* of point M .

Moreover, point M' will be called the *inverse* of M with respect to line u if M' is symmetrical to M with respect to line u . In the following exposition, as a rule, we shall not distinguish between inversion relative to a circle and inversion relative to a line, treating a line as a circle with an infinite radius. Proofs of theorems concerning inversions will be usually carried out under the assumption that the circle of inversion is the usual circle. The special case where the circle of inversion has an infinitely large radius (that is, is a line) sometimes requires additional but trivial arguments. These the reader can reproduce himself.

The following properties of inversions are obvious.

(1) If M' is the inverse of point M , then M is the inverse of point M' . Thus inversion coincides with its inverse mapping.

(2) Under inversion the exterior region of circle k in the plane is mapped onto the interior region and vice versa.*

* If the radius of circle k is infinitely large, that is, it is a line, then any of the two half planes defined by this line can be treated as the interior region and then the second as the exterior.

(3) *Every point of circle k coincides with its inverse.*

Other properties of inversion involve a little computation. In a plane let us introduce an orthogonal Cartesian system of coordinates and with each point M associate the complex number $z = x + iy$, where x and y are the coordinates of M . As usual, a bar over z will denote the complex conjugate of the given number: $\bar{z} = x - iy$. Evidently, point M is determined by prescribing any one of the two numbers, z or $\bar{z}$.

Let the origin be the centre of the circle of inversion. Then if two points determined by z and z' are inverses of each other, (*) yields a relation between z and z' :

$$\bar{z}z' = r^2$$

This gives an analytic representation of inversion:

$$z' = r^2/\bar{z}$$

or, in terms of coordinates,

$$x' = r^2 \frac{x}{x^2 + y^2}, \quad y' = r^2 \frac{y}{x^2 + y^2}$$

Using these formulas the so-called *circular property of inversion* may be proved: if point z describes a circle or a line, its inverse z' describes either a circle or a line.

Treating a line as a circle of infinitely large radius, we can express the last result as follows:

(4) *The inverse of a circle is a circle.*

To prove this, consider any circle

$$A(x^2 + y^2) + Bx + Cy + D = 0$$

Replacing the current coordinates x, y by

$$x = r^2 \frac{x'}{x'^2 + y'^2}, \quad y = r^2 \frac{y'}{x'^2 + y'^2}$$

we have

$$Ar^4 + Br^2x' + Cr^2y' + D(x'^2 + y'^2) = 0$$

Thus the coordinates of points which are the inverse of points of a circle also satisfy the equation of a circle (or of a line if $D = 0$). Thus the statement is proved.

A special role in our discussion is played by special mappings which are products of inversions.

Let us consider one such mapping which transforms any point z into a point z' . It is easy to show that if this mapping is the product of an even number of inversions, then z' is given in terms of z by the

formula

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta} \quad (\text{I})$$

where α , β , γ , δ are complex constants. On the other hand, if the given mapping consists of an odd number of inversions, then z' is given by

$$z' = \frac{\overline{\alpha z + \beta}}{\overline{\gamma z + \delta}} \quad (\text{II})$$

We shall first show that inversion with respect to a circle with a centre a and a radius r is analytically given by the formula (II). To do this, we introduce a new coordinate system with origin at a and axes parallel to the old axes. If M and M' are two points inverse with respect to the given circle, and Z and Z' are corresponding complex numbers in the new system, then

$$Z' = r^2 / \overline{Z}$$

Let z and z' be complex numbers corresponding to these points in the old system; evidently, $z = Z + a$ and $z' = Z' + a$. Substituting these into the last relation, we find

$$z' - a = \frac{r^2}{\overline{z - a}}$$

whence

$$z' = \frac{a\overline{z} + (r^2 - a\overline{a})}{\overline{z} - \overline{a}}$$

or, if we set $a = \alpha$, $r^2 - a\overline{a} = \beta$, $1 = \gamma$, $-\overline{a} = \delta$, we get

$$z' = \frac{\alpha\overline{z} + \beta}{\gamma\overline{z} + \delta}$$

The above computation was carried out with the assumption that the inversion is with respect to an ordinary circle. But it is not difficult to obtain a relation between z and z' if the inversion is performed with respect to a line. The inversion with respect to the real axis is given by the equation $z' = \overline{z}$; consequently, inversion relative to a line passing through the origin is analytically given by the equation $e^{i\varphi}z' = \overline{(e^{i\varphi}z)}$, or $z' = e^{-2i\varphi}\overline{z}$. This will yield, with the aid of translation, the relationship between z and z' when the line of inversion has any position, namely

$$z' = e^{-2i\varphi}\overline{z} + \text{constant}$$

This is the same as (II) with $\gamma = 0$.

Thus the equation $z' = \frac{\alpha\bar{z} + \beta}{\gamma\bar{z} + \delta}$ with a proper choice of constants $\alpha, \beta, \gamma, \delta$ defines any inversion with respect to either an ordinary circle or a circle with an infinitely large radius.

Suppose that two inversions are successively performed with respect to any circle. If the first inversion maps z into z' and the second z' into z'' , then

$$z' = \frac{\alpha_1\bar{z} + \beta_1}{\gamma_1\bar{z} + \delta_1}$$

and

$$z'' = \frac{\alpha_2\overline{z'} + \beta_2}{\gamma_2\overline{z'} + \delta_2}$$

The first of these two relations implies that

$$\overline{z'} = \frac{\bar{\alpha}_1 z + \bar{\beta}_1}{\bar{\gamma}_1 z + \bar{\delta}_1}$$

Substitution of this expression into the second equation, after minor computations and introduction of proper notation, yields

$$z'' = \frac{\alpha z + \beta}{\gamma z + \delta}$$

that is, a relationship of type (I). If another inversion is performed which maps z'' into z''' , then evidently, the relationship between z''' and z will have form (II); the next inversion of z''' will again yield (I), and so on.

We shall now prove the properties of products of inversion to be used later.

(5) *If a mapping consisting of an even number of inversion leaves three points of the plane fixed, then all the points of the plane retain their positions under this mapping, the mapping is thus an identity.*

As noted above, mapping of this form transforming z into z' is characterized by the equation

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta}$$

All the fixed points of this mapping are given by the equation $z' = z$, that is, by

$$z = \frac{\alpha z + \beta}{\gamma z + \delta}, \quad \text{or} \quad \gamma z^2 + (\delta - \alpha)z - \beta = 0$$

According to the condition, this equation must have three solutions. This is possible only if it reduces to an identity, that is, if

$$\gamma = 0, \quad \delta - \alpha = 0, \quad \beta = 0$$

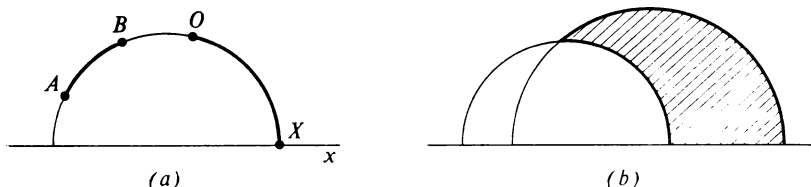


Figure 68

Hence

$$z' = (\alpha/\delta) z$$

Clearly, $\alpha \neq 0$ (if $\alpha = 0$, then every point z is mapped into the same point $z' = 0$, which is impossible for the mapping which is a product of inversions because each inversion maps distinct points into distinct points. Since $\delta - \alpha = 0$, at $\alpha \neq 0$, we have $z' = z$. ■

(6) *If a mapping consisting of the product of an odd number of inversions leaves three points of the plane fixed, then this mapping is an inversion with respect to a circle through these three points.*

Assume that $z' = f(z)$ is the given mapping. If $z'' = \varphi(z')$ is an inversion with respect to the abovementioned circle, then $z'' = \varphi(f(z))$ is a mapping composed of an even number of inversions and it leaves the same three points fixed as the given mapping $z' = f(z)$. Then, according to the preceding result, $z'' = \varphi(f(z))$ must be an identity, $z'' = z$. Thus $\varphi(z') = z$ and accordingly, z and z' the inverse of each other with respect to the circle passing through the three points in question. ■

In conclusion, we formulate without proof another proposition concerning inversions.

(7) *If two circles intersect, then an inversion preserves the angle between these circles.*

The invariance of an angle under an inversion is proved in the elementary theory of conformal mapping.

Now we may return to the construction of a model for non-Euclidean geometry.

§ 52. According to the definition of “betweenness” given in § 50, the order of points on a non-Euclidean line coincides with the order of points on the Euclidean semicircle representing this line in the upper half plane. Therefore, a non-Euclidean segment AB is represented by an arc of the semicircle with endpoints A and B ; the non-Euclidean half line emanating from the point O is represented by arc OX whose endpoint X lies on the line x (Fig. 68a). Of course, X is not considered as a point of the non-Euclidean half line.

A *non-Euclidean angle* is defined as the totality of two non-Euclidean half lines emanating from one point (Fig. 68b).

We shall now define the congruence of segments and angles in our model of non-Euclidean geometry.

We shall make considerable use of inversions, and consider only those inversions which are performed with respect to circles orthogonal to line x . Evidently, such inversions map points of the upper half plane into points of the same half plane. Thus in performing inversions of figures of the upper half plane we do not go beyond its limits.

A non-Euclidean segment AB is *congruent* to another non-Euclidean segment $A'B'$ if there is a chain of inversions such that their product maps a Euclidean circular arc AB into a circular arc $A'B'$.

Similarly, a non-Euclidean angle $\angle(h, k)$ is congruent to another non-Euclidean angle $\angle(h', k')$ if there is a chain of inversions such that their product maps the sides of the first angle onto the sides of the second angle.*

Proposition (7) of § 51 implies that angles congruent in the sense of the present definition are equal among themselves also in the sense of Euclidean geometry regarding curvilinear angles. In contrast, circular arcs representing non-Euclidean congruent segments are not congruent in the sense of Euclidean geometry since inversions, while preserving the magnitudes of angles, affect linear magnitudes of figures.

In our model of non-Euclidean geometry inversions with respect to circles orthogonal to line x are some sort of congruent displacements. Let us examine their peculiarities more closely.

Consider a semicircle orthogonal to x in the upper half plane. According to Proposition (4) of § 51 an inversion maps this semicircle into a circular arc (also in the upper half plane) and the line x is mapped onto itself. Since an inversion preserves the magnitude of angles, the circular arc obtained by inversion of the given semicircle must be orthogonal to line x and hence must also be a semicircle. Thus the inversion in question always maps a semicircle orthogonal to x in the upper half plane into the same type of semicircle. This is a very important feature because semicircles orthogonal to x in the upper half plane are lines in our model of non-Euclidean geometry.

Let AB be a circular arc representing a non-Euclidean segment (Fig. 69). Denote by S the point of intersection of Euclidean line AB with line x (assuming they intersect) and draw from S a tangent SC to the arc AB . By a known theorem of Euclidean geometry, $SA \times SB = SC^2$. Therefore, if u denotes a semicircle with the

* Congruence relations of segments and angles are mutual; this is a consequence of the fact that inverse mappings of inversions are also inversions.

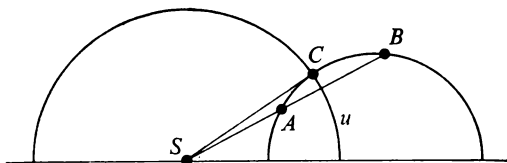


Figure 69

centre S and radius SC , inversion with respect to u maps A into B and B into A but point C remains unchanged. This means that arc AB is mapped onto itself, and its part AC is mapped onto BC and BC onto AC . Since both of them are inverse of each other, arcs AC and CB represent non-Euclidean congruent segments, and C is the non-Euclidean midpoint of arc AB . Note also that arc AB is orthogonal to semicircle u . Thus u represents a perpendicular at the midpoint of the non-Euclidean segment AB . In other words, points A and B are symmetric, in the non-Euclidean sense, with respect to a non-Euclidean line represented by u . We may thus conclude that from the non-Euclidean point of view an inversion is nothing but symmetry with respect to a line.

The aforesaid construction assumed the existence of the point S . If the Euclidean line AB does not intersect line x , then S should be understood as a point at infinity, the tangent to arc AB as a line parallel to x , while u should be replaced by a half line. In this case inversion reduces to the ordinary symmetry with respect to the Euclidean perpendicular to line x drawn from the Euclidean midpoint C of the arc AB . This makes clear what is to be understood by the definition of congruence of images given earlier: an image A is congruent to another image A' if A' may be obtained from A by means of a chain of reflections understood in the aforementioned (non-Euclidean) sense.

Our aim now is to show that the congruence relations thus introduced satisfy Axioms III.1-III.5.

These axioms will be dealt with successively. Axiom III.1 requires that a given segment of a given line be congruent to a segment that originates from any point on another line and lies on a given side of this point. This axiom is satisfied in our model. Consider the two non-Euclidean lines, a and a^* , and take segment AB on one of them and a point A^* on the second (Fig. 70). Moreover, choose one of the two half lines starting with the point A^* on a^* . As shown above, draw a (non-Euclidean) perpendicular n at the midpoint of segment AA^* . By means of a (non-Euclidean) reflection in this perpendicular, line a may be mapped onto another line a' ; then point A is mapped into A^* and segment AB of line a into segment A^*B' of line a' . Now draw a (non-Euclidean) bisector b of the angle formed by two (non-

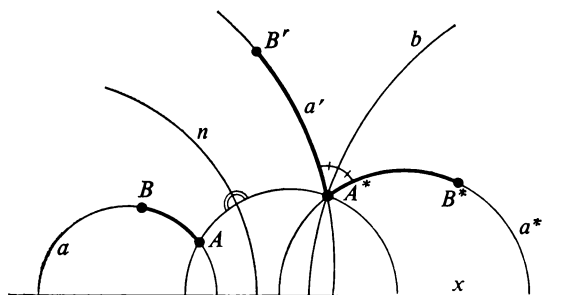


Figure 70

Euclidean) half lines one of which originates from A^* towards B' and the other is the chosen half line on line a^* . The reflection (understood in the non-Euclidean sense) in b maps (non-Euclidean) line a' into (non-Euclidean) line a^* and segment A^*B' of line a' into a segment A^*B^* . Hence there is a point B^* on (non-Euclidean) line a^* on the given side of its point A^* such that segment A^*B^* is obtained from segment AB by means of two (non-Euclidean) reflections; accordingly, $AB \equiv A^*B^*$ in the accepted sense. ■

Axiom III.1 also requires that among the points of line a^* lying on a given side of A^* there is only one point which together with A^* determines a segment congruent to the segment AB . We shall show that this requirement is fulfilled for our definition of congruence.

Suppose that on a given side of point A^* there are two distinct points B_1^* and B_2^* on (non-Euclidean) line a^* such that $AB \equiv A^*B_1^*$ and $AB \equiv A^*B_2^*$. This means that there is a chain of inversions whose product maps the circular arc AB into the circular arc $A^*B_1^*$ and another chain of inversions whose product maps arc AB into the circular arc $A^*B_2^*$. We denote by X_2 the point at which the extension of arc AB in the direction from A to B meets line x , and by X_1 we denote the point of intersection of x with arc AB extended in the opposite direction. Let X_2^* and X_1^* denote analogously determined endpoints of the semicircle representing the non-Euclidean line a^* . Evidently, the product of each of the two chains of inversions maps X_1 into X_1^* and X_2 into X_2^* . Suppose that inversion in the first chain is performed in the reverse order and then inversions of the second chain are performed. This results in a certain mapping denoted by f . Evidently, under f point X_1^* first coincides with X_1 and then goes back to X_1^* , which means that X_1^* is a fixed point of f . Similarly, the points A^* and X_2^* are also fixed points of f . As regards B_1^* , f maps it into B_2^* . Thus f has three fixed points X_1^* , A^* and X_2^* , which, in view of Propositions (5) and (6) of § 51, implies that f is either the identity mapping or is inversion with respect to the circle through

points X_1^* , A^* , X_2^* , B_1^* and B_2^* . In either case, all the points of this circle are fixed points of f ; accordingly, B_1^* and B_2^* must be distinct. This proves that the operation of marking off the congruent segment is unique.

Finally, Axiom III.1 requires that segment AB be congruent to segment BA . In order to show that this requirement is fulfilled in our model of non-Euclidean geometry, it suffices to perform a non-Euclidean reflection in the midpoint of segment AB .

All the requirements of Axiom III.1 are thus satisfied.

Let us now examine Axiom III.2, according to which if segments $A'B'$ and $A''B''$ are congruent to segment AB , then $A'B'$ is congruent to $A''B''$. This requirement is evidently fulfilled in our model of non-Euclidean geometry. In fact, relations $A'B' \equiv AB$ and $A''B'' \equiv AB$ imply that there is a chain of non-Euclidean reflections mapping $A'B'$ into AB and another chain of non-Euclidean reflections mapping $A''B''$ also into AB . If the reflections of the first chain are performed and then reflections of the second chain in the reverse order, then $A'B'$ is mapped into $A''B''$, implying the congruence of these segments.

Consider now Axiom III.3. Let AB and $A'B'$ be non-Euclidean segments, C an interior point of AB and C' an interior point of $A'B'$. We have to show that in our definition of congruence, $AC \equiv A'C'$ and $CB \equiv C'B'$ imply $AB \equiv A'B'$. Since $AC \equiv A'C'$, there is a chain of non-Euclidean reflections whose product maps AC into $A'C'$. Simultaneously, B is mapped into B^* on the line $A'B'C'$, where B^* lies on the same side of C' as B' . Segments $C'B'$ and $C'B^*$ lie on the same side of point C' and, being congruent to segment CB ($C'B'$ is congruent to CB by assumption and $C'B^*$ is congruent to CB by construction), must be congruent to each other (by Axiom III.2, which has already been shown to be valid). But then, by Axiom III.1, points B^* and B' cannot be distinct. Thus the product of the aforementioned non-Euclidean reflections maps AB into $A'B'$, implying $AB \equiv A'B'$.

To verify the validity of Axiom III.4 is also a simple matter. This axiom requires that a given angle be congruent to an angle that can be adjoined on any side of a half line and that this operation be unique.

That this operation is possible and unique is established in the same way as was done for Axiom III.1. Namely, let $\angle(h, k)$ be a non-Euclidean angle with vertex O and let h' be a non-Euclidean half line with vertex O' . First, by means of a non-Euclidean reflection in the perpendicular at the midpoint of segment OO' , the angle $\angle(h, k)$ is mapped into angle $\angle(h'', k'')$ whose vertex is at O' . Next, a non-Euclidean reflection in the bisector $\angle(h', h'')$ maps $\angle(h'', k'')$ into $\angle(h', k')$; this angle is congruent to $\angle(h, k)$ and is adjoined to the half line h' on a given side. Now

we must prove that the operation of adjoining an angle to a given half line on a given side is unique. Let $\angle (h, k)$ be mapped into $\angle (h', k'_1)$ by a chain of non-Euclidean reflections and let $\angle (h, k)$ be mapped into $\angle (h', k'_2)$ by another chain of non-Euclidean reflections. Let f denote the mapping which is the product of reflections of the first chain performed in reverse order and reflections of the second chain. Evidently, f maps $\angle (h', k'_1)$ into $\angle (h', k'_2)$. But noting that non-Euclidean reflections are inversions and applying Propositions (5) and (6) of § 51 it may be shown, in exactly the same way as was done in the case of Axiom III.1, that f is either the identity mapping or a non-Euclidean reflection in the half line h' . Consequently, $\angle (h', k'_1)$ and $\angle (h', k'_2)$ either coincide or are reflections of each other (in the non-Euclidean sense) in h' . ■

Axiom III.4 also requires that every angle $\angle (h, k)$ be congruent to itself, that is, $\angle (h, k) \equiv \angle (h, k)$ and $\angle (h, k) \equiv \angle (k, h)$. But the first relation is obvious* and the second relation may be proved by performing a non-Euclidean reflection of the angle in its bisector.

Finally, Axiom III.5 in our model can be shown to be valid if one proceeds with the same arguments as used in school textbooks on elementary geometry in proving the first theorem on equality of triangles, treating a motion as the result of a certain number of non-Euclidean reflections.

Thus we find that the congruence relation in the constructed system of objects satisfy the conditions of all the axioms of the third group.

One can now immediately conclude that axioms of continuity IV.1 and IV.2 are also satisfied in this system of objects. In fact, as noted in § 50, Dedekind's principle is realized on non-Euclidean lines. And, in view of Theorem 41 of § 23, Dedekind's principle together with Axioms I-III implies Axioms IV.1 and IV.2.

Thus our system of objects meets the requirements of all the axioms of absolute plane geometry: Axioms I.1-I.3 and the axioms of Groups II-IV. But then for this system there must hold either the Euclidean or the Lobachevskian theory of parallels. As will be shown now, the second theory holds for our system.

Consider, in the upper half plane, a semicircle a orthogonal to a line x . Let A be any point of the upper half plane not on this circle (Fig. 71). We can easily see that there pass infinitely many distinct semicircles through A which are orthogonal to x and do not intersect semicircle a . In our agreed terminology, this fact may be stated thus: through any non-Euclidean point there pass infinitely many distinct non-Euclidean lines not intersecting the given line. This implies that in the system of objects under consideration Lobachevsky's postulate holds; this system is therefore a model of Lobachevsky's geometry.

* Since the identity mapping may be regarded as a double inversion.

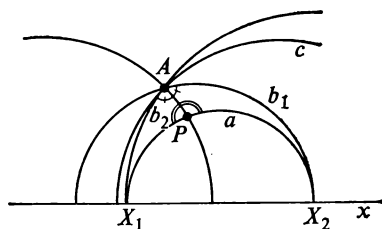


Figure 71

kian geometry the construction of which was our aim. Using this model, each proposition of Lobachevskian planimetry can be fully interpreted on the Euclidean plane. To do so, the words "point", "line", "congruent to", etc. appearing in the formulation of propositions must be understood in the agreed sense, that is, the word "point" means a Euclidean point of the upper half plane, "line" a Euclidean semicircle or a half line orthogonal to the boundary of the semicircle; those images are called congruent which may be mapped into each other by means of successive inversions, and so on. Thus to every Lobachevskian theorem there corresponds a fully defined Euclidean theorem. Therefore, if there were any inconsistency in Lobachevskian geometry, it would also appear in Euclidean geometry.

Thus we find that consistency of Lobachevskian geometry is a consequence of consistency of Euclidean geometry. At the same time we have shown that Euclid's parallel postulate cannot be derived from the premises of absolute geometry.

Indeed, in the model constructed by Poincaré all the axioms of absolute geometry are realized, but instead of the Euclidean parallel postulate the Lobachevskian postulate holds; this means that the Euclidean postulate is not a logical consequence of these axioms.

§ 53. It is interesting to see how the concrete facts of Lobachevskian geometry are interpreted on the Euclidean half plane.

Let us examine Fig. 71. Here a point A and a non-Euclidean line in the form of a semicircle a orthogonal to line x have been drawn. Non-Euclidean lines through A which do not intersect the given line have been represented by semicircles which are orthogonal to x and pass through A but do not intersect the semicircle a . Among these non-Euclidean lines two boundary lines must exist which are called parallel to the given line in two directions. In Fig. 71, parallel lines are represented by semicircles b_1 and b_2 that are tangent to a at endpoints X_1 and X_2 of x . The Euclidean points of x are not taken as non-Euclidean objects; therefore non-Euclidean lines represented by semicircles b_1 and b_2 should not be viewed as intersecting with line a . That they are boundary lines is seen directly.

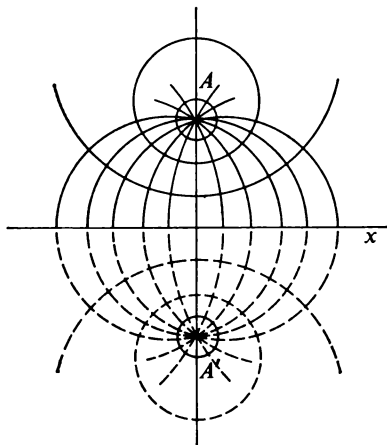


Figure 72

Draw through A a semicircle which is orthogonal to x and intersects semicircle a orthogonally at P . The arc AP evidently represents the non-Euclidean perpendicular to non-Euclidean line a , and the angle this perpendicular forms with arc b_1 is nothing but the angle of parallelism for segment AP .

Totally trivial for Lobachevskian geometry is the fact that perpendicular AP is a bisector of the angle between non-Euclidean lines b_1 and b_2 . In Euclidean geometry the fact that the angles formed by AP with arcs b_1 and b_2 are equal is not obvious, but the need to prove such a Euclidean theorem does not arise. Indeed, since the system of objects in Poincaré's model satisfies all the axioms of Lobachevskian geometry, all the statements following from these axioms, including the one mentioned just now, must be true. This, in particular, shows a novel way of proving certain Euclidean theorems by means of non-Euclidean geometry.

To cite another example, consider the following Euclidean theorem, which may be shown to be valid without any detailed proof: if a triangle is formed by circular arcs which intersect, if extended, a given line at right angles, the sum of the interior angles of this triangle is less than two right angles. This theorem is obtained from the corresponding theorem in Lobachevskian geometry by means of Poincaré's interpretation.

Let us examine non-Euclidean circles, equidistants and oricycles in Poincaré's model. These curves are orthogonal trajectories of elliptic, hyperbolic and parabolic pencils composed of non-Euclidean lines (see the concluding portion of § 39).

In Fig. 72 there is a pencil of Euclidean circles having two nodal points A and A' , of which A is in the upper half plane and A' is

symmetrically situated in the lower half plane. The orthogonal trajectories of this pencil are also circles, which also form a pencil but without nodal points and having limiting points A and A' (the proof of this fact is omitted). Clearly, the upper halves of circles of the first pencil represent non-Euclidean lines passing through point A and therefore form an elliptic pencil and the circles of the second pencil, orthogonal to circles of the first, lying in the upper half plane represent non-Euclidean circles with a common centre A .

Figure 73 shows semicircle a orthogonal to line x and a pencil of circles orthogonal to a with limiting points X_1 and X_2 . The upper halves of these circles represent non-Euclidean lines with a common perpendicular a ; the totality of such lines forms a hyperbolic pencil with base a . Any circle through X_1 and X_2 is an orthogonal trajectory of this pencil and hence the upper arc of such a pencil represents an equidistant whose base is the non-Euclidean line a . The circular arcs A_1B_1 , A_2B_2 , A_3B_3 , . . . , etc. represent altitudes of equidistant b which are congruent in the non-Euclidean sense.

Figure 74 shows a pencil of circles with coincident nodal points; the upper halves of these circles represent non-Euclidean lines parallel to each other in one direction and therefore form a rectilinear non-Euclidean parabolic pencil. Its orthogonal trajectories, from the non-Euclidean viewpoint, are oricycles, while as objects of the Euclidean plane they are circles which are tangent to each other and to the line x at the nodal point.

Thus a circular arc lying in the upper half plane depicts a non-Euclidean line if this arc is supported by line x and forms with it a right angle, an equidistant if it is supported by line x and forms with it an angle other than a right angle, an oricycle if its endpoints are joined and the arc is tangent to line x at the point of intersection,

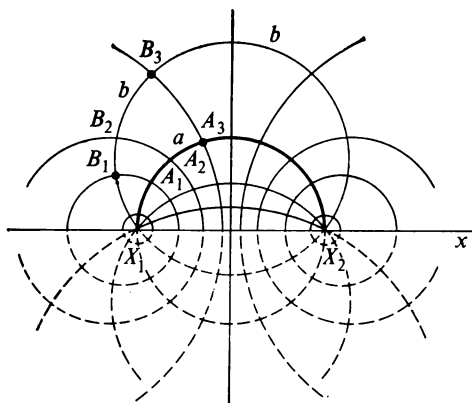


Figure 73

and, finally, a non-Euclidean circle if it is a full Euclidean circle in the upper half plane.

§ 54. The aforementioned interpretation of non-Euclidean geometry is by no means the only one possible. Apart from this interpretation, there are infinitely many other interpretations. What follows is another example of how non-Euclidean geometry may be interpreted on the Euclidean plane.

Consider a circle K in the Euclidean plane. By non-Euclidean points we mean points of the Euclidean plane lying inside K , and by non-Euclidean lines we mean the portion of Euclidean circles lying inside K (including the diameters) orthogonal to the given circle. The mutual relationships between the geometric elements and the "betweenness" are understood in the Euclidean sense.

Two non-Euclidean images will be considered to be reflections of each other in the non-Euclidean sense if their Euclidean images inside K can be mapped into each other by means of inversion with respect to a circle orthogonal to K . Two non-Euclidean images will be said to be congruent if they can be mapped into each other by means of a chain of non-Euclidean reflections.

Proceeding in the same way as in §§ 50-52, we may show that with these definitions of geometric elements and their mutual relationships all the requirements of the axioms of absolute geometry are satisfied. Then it is an easy matter to settle which of the parallel theories is realized in the system of non-Euclidean lines inside the circle K . Let a be a circular arc orthogonal to circle K and A a point inside K but not on arc a (Fig. 75). By means of elementary geometry, it may be shown that through A there pass infinitely many circular arcs orthogonal to K which do not intersect a . This

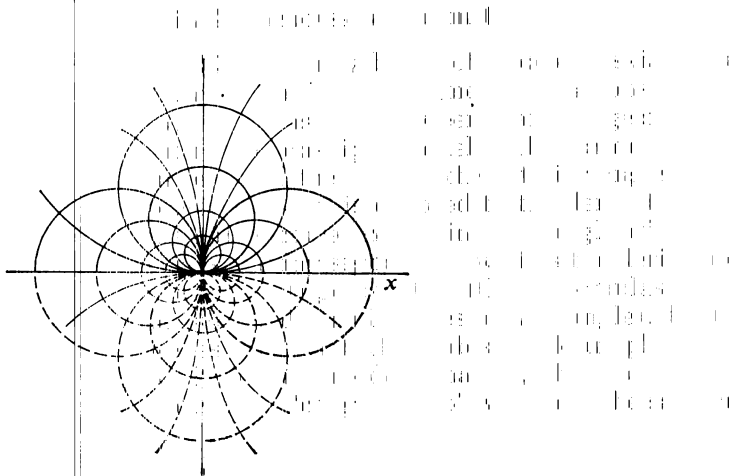


Figure 74

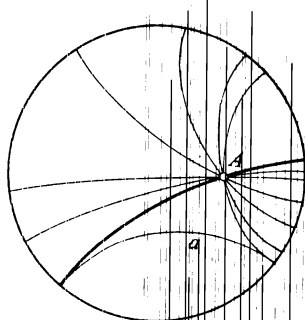


Figure 75

means that in the sense of those relationships which have been established for non-Euclidean images inside K , the Lobachevskian parallel postulate holds for objects of this system. Thus a new interpretation of Lobachevskian plane geometry on the Euclidean plane is obtained.

Each proposition of Lobachevskian geometry stated in an abstract form may be interpreted on the Euclidean half plane or inside a Euclidean circle; this will result in a certain theorem of Euclidean geometry whose concrete meaning will depend on the manner of interpretation. The possibility of obtaining Euclidean theorems in this manner from the Lobachevskian abstract logical scheme is explored in the geometric theory of the functions of a complex variable, where a close connection is also established between the two aforementioned interpretations of Lobachevskian geometry and where general principles are laid down for various other interpretations.

3.10. The basic metric relationships in Lobachevskian geometry

§ 55. The novelities of Lobachevskian geometry become especially clear when its metric relationships are examined, that is, the relationships between various geometrical quantities. One of such relationships, namely, the formula for the area of a triangle in terms of the sum of the interior angles, was treated in § 48. The present section is devoted to the derivation of Lobachevsky's fundamental formula expressing the angle of parallelism in terms of the corresponding segment as well as the derivation of formulas of Lobachevskian trigonometry (these formulas establish the relationship between sides and angles of a triangle). In deriving these formulas, we assume that the Lobachevskian plane is realized in the form of Poincare's model, that is, the words "point", "line", "lies on", "between", "congruent to" will have the same meaning as in §§ 50-52.

Such a derivation of Lobachevsky's basic formulas is simple and clear; what is more, it nicely brings out the connection between Lobachevskian geometry and the theory of a complex variable. Of course, one cannot assert that these formulas are generally valid in any Lobachevskian geometry, that is, they hold for any of its models.

In Chapter 8, the same formulas will be obtained by means of axioms and will not be connected with any model. This will show that these formulas are valid for any model of Lobachevskian geometry. This derivation of fundamental formulas of Lobachevskian geometry in Chapter 8 is also very simple but is based on certain results of projective geometry. These propositions are formulated in Chapter 8; therefore, a reader who is prepared to accept them may skip the chapters devoted to projective geometry and go directly to the abstract derivation of Lobachevsky's formulas (see §§ 216-221, 229-232).

§ 56. First we shall give a short account of the invariants of linear-fractional transformations of a complex variable. As usual, $z = x + iy$ will denote a point with rectangular Cartesian coordinates x and y ; the phrases "number z " and "point z " will be used synonymously.

Consider the transformation

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta} \quad (1)$$

where z is a complex variable, and α , β , γ , δ are constants (generally speaking, complex). The transformation of z into z' according to the formula of type (1) is called *linear-functional** (we have already had occasion to use them in § 51). It is clear that either γ or δ in (1) should be nonzero.

The quantity $\Delta = \alpha\delta - \beta\gamma$ is known as the *determinant* of the linear-fractional transformation (1). It may be easily verified that if $\Delta = 0$, then according to formula (1) all points z (with the condition $\gamma z + \delta \neq 0$) have as their corresponding points the same point z' . Indeed, if $\Delta = 0$, the numbers α and β are proportional to γ and δ , that is, $\alpha = k\gamma$, $\beta = k\delta$ and therefore

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta} = \frac{k(\gamma z + \delta)}{\gamma z + \delta} = k$$

On the other hand, if $\Delta \neq 0$, distinct points z_1 and z_2 are mapped by formula (1) into distinct points z'_1 and z'_2 . Indeed, we have

$$z'_1 - z'_2 = \frac{\alpha z_1 + \beta}{\gamma z_1 + \delta} - \frac{\alpha z_2 + \beta}{\gamma z_2 + \delta} = \frac{\alpha\delta - \beta\gamma}{(\gamma z_1 + \delta)(\gamma z_2 + \delta)} (z_1 - z_2)$$

* The terms bilinear transformation or sometimes even linear transformation are also used.

that is,

$$z'_1 - z'_2 = \frac{\Delta}{(\gamma z_1 + \delta)(\gamma z_2 + \delta)} (z_1 - z_2) \quad (2)$$

Thus, if $\Delta \neq 0$ and $z_1 \neq z_2$, then $z'_1 \neq z'_2$.

If $\Delta = 0$, the linear-fractional transformation is called *degenerate*, and *nondegenerate* if $\Delta \neq 0$. According to the previous result, a degenerate transformation maps all points of the plane into only one point, while a nondegenerate transformation maps distinct points into distinct points. In either case, the point z satisfying $\gamma z + \delta = 0$ must be dropped from consideration; there is no point corresponding to it.

Only nondegenerate linear-fractional transformations will be discussed below. For us it is important that every nondegenerate linear-fractional transformation of form (1) has an inverse transformation

$$z = \frac{\delta z' - \beta}{-\gamma z' + \alpha} \quad (3)$$

which is also linear-fractional and nondegenerate (transformation (3) is nondegenerate because its determinant $\Delta' = \alpha\delta - \beta\gamma = \Delta \neq 0$).

The presence of the singular point $z = -\delta/\gamma$ for which formula (1) fails to be valid causes some complications in formulating propositions regarding linear-fractional transformations. To simplify the matter, we adjoin to the complex plane a new object, called the *point at infinity* and denoted by the symbol ∞ ; we assume that in the case of a nondegenerate transformation of form (1) the point $z = -\delta/\gamma$ is mapped into the point at infinity. The point at infinity is taken as the image of a singular point of every nondegenerate linear-fractional transformation. In particular, under transformation (3) the point at infinity is the image of point $z' = \alpha/\gamma$. Since transformations (1) and (3) are the inverse of each other, the point at infinity should be treated as the *preimage* of the point $z' = \alpha/\gamma$ under transformation (1). Thus, according to our assumption, nondegenerate transformation (1) maps point $z = -\delta/\gamma$ into $z' = \infty$ and point $z = \infty$ into $z' = \alpha/\gamma$.

Finally, note that if $\gamma = 0$, there is no singular point and every point of the plane has an (ordinary) image. We shall agree that through such transformations the point at infinity is mapped into itself.

Let u, v, s, t be four distinct points, and assume that they are nonsingular (that is, there is no point at infinity among them). Then the quantity $(uvst)$ defined by

$$(uvst) = \frac{u-s}{u-t} \div \frac{v-s}{v-t} \quad (3')$$

is called the *cross ratio* of points u, v, s, t taken in this order. A cross ratio of the same points but taken in another order may have a differ-

ent value; for example, if $(uvst) = \lambda$, then

$$(vust) = \lambda^{(-1)}, \quad (uwts) = \lambda^{(-1)} \quad (4)$$

If the point at infinity is one of the four points, the cross ratio is defined by one of the following formulas:

$$\left. \begin{aligned} (wvs\infty) &= \frac{u-s}{v-s}, & (uv\infty t) &= \frac{v-t}{u-t} \\ (u\infty st) &= \frac{u-s}{u-t}, & (\infty vst) &= \frac{v-t}{v-s} \end{aligned} \right\} \quad (5)$$

Note that the first of these formulas is obtained by letting $t \rightarrow \infty$ in (3'), the second by letting $s \rightarrow \infty$, and so on.

The cross ratio of four points is invariant under nondegenerate linear-fractional transformations; this means that if by a nondegenerate linear-fractional transformation points u, v, s, t are mapped into points u', v', s', t' , then

$$(u'v's't') = (uvst)$$

We first prove this result for the case when neither the given points nor their images contain a point at infinity. Suppose that the transformation mapping u, v, s, t into u', v', s', t' is given by formula (1). Then formula (2) gives

$$u' - s' = \frac{\Delta}{(\gamma u + \delta)(\gamma s + \delta)} (u - s), \quad u' - t' = \frac{\Delta}{(\gamma u + \delta)(\gamma t + \delta)} (u - t)$$

whence

$$\frac{u' - s'}{u' - t'} = \frac{u - s}{u - t} \frac{\gamma t + \delta}{\gamma s + \delta}$$

Similarly,

$$\frac{v' - s'}{v' - t'} = \frac{v - s}{v - t} \frac{\gamma t + \delta}{\gamma s + \delta}$$

Accordingly,

$$\frac{u' - s'}{u' - t'} \div \frac{v' - s'}{v' - t'} = \frac{u - s}{u - t} \div \frac{v - s}{v - t}$$

that is, $(u'v's't') = (uvst)$.

Suppose now that all the points u, v, s, t are ordinary points and one of the points u', v', s', t' is at infinity, say $t' = \infty$. This means that $\gamma t + \delta = 0$. Then

$$\begin{aligned} (u'v's't') &= (u'v's'\infty) = \frac{u' - s'}{v' - s'} = \frac{u - s}{v - s} \frac{\gamma v + \delta}{\gamma u + \delta} \\ &= \frac{u - s}{v - s} \frac{v + (\delta/\gamma)}{u + (\delta/\gamma)} = \frac{u - s}{v - s} \cdot \frac{v - t}{u - t} = \frac{u - s}{u - t} \div \frac{v - s}{v - t} = (uvst) \end{aligned}$$

The case where one of the points u, v, s, t is at infinity and all the points u', v', s', t' are ordinary reduces to the preceding case.

In fact, using the inverse of the given transformation, the last result yields $(uvst) = (u'v's't')$.

We still have to consider the case where one of the points u, v, s, t is at infinity and it is mapped into itself; this case occurs if $\gamma = 0$ in the transformation given by formula (1).

Suppose that $t = \infty$ and $t' = \infty$. Then

$$\begin{aligned}(u'v's't') &= (u'v's'\infty) = \frac{u'-s'}{v'-s'} = \frac{u-s}{v-s} \frac{\gamma v + \delta}{\gamma u + \delta} \\ &= \frac{u-s}{v-s} \frac{\delta}{\delta} = \frac{u-s}{v-s} = (uvs\infty) = (uvst)\end{aligned}$$

Thus in all cases $(uvst) = (u'v's't')$ and the result is proved.

§ 57. We shall require one more transformation mapping z into z' which is given by

$$z' = \frac{\alpha \bar{z} + \beta}{\gamma \bar{z} + \delta} \quad (6)$$

and is known as a linear-fractional transformation of the second kind (recall that $\bar{z}$ denotes the complex conjugate of z). Such a transformation was used in § 51.

A transformation of form (6) is called *degenerate* if $\Delta = \alpha\delta - \beta\gamma = 0$ and *nondegenerate* if $\Delta \neq 0$; a degenerate transformation maps all the points into the same one point, while a nondegenerate transformation maps distinct points into distinct points (these results are proved in the same way as analogous results for a transformation of the first kind). Among transformations of form (6) only the nondegenerate transformations will be considered below.

Let u, v, s, t be any four distinct points and u', v', s', t' their images under a nondegenerate transformation of form (6). Then the cross ratio of points u', v', s', t' is a complex conjugate of the cross ratio of u, v, s, t . This is denoted thus:

$$(u'v's't') = \overline{(uvst)}$$

To establish this result, present (6) in the form of a product of two transformations

$$z' = \frac{\alpha z'' + \beta}{\gamma z'' + \delta} \quad (7)$$

and

$$z'' = \bar{z} \quad (8)$$

The point at infinity is mapped into itself by transformation (8).

Note that if all the numbers occurring in cross ratio are replaced by their complex conjugates, the cross ratio itself is replaced by its conjugate. Therefore, denoting the images of u, v, s, t by $u'', v'',$

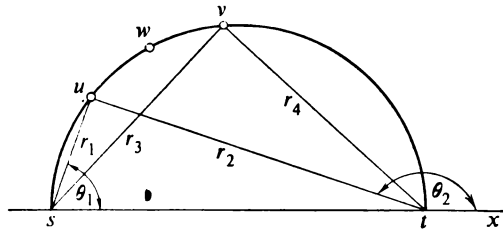


Figure 76

s'', t'' under transformation (8), we have

$$(u''v''s''t'') = (\overline{uvst})$$

Further, since (7) is a linear-fractional transformation of the first kind,

$$(u'v's't') = (u''v''s''t'')$$

The last two relations yield the desired result

$$(u'v's't') = (\overline{uvst})$$

§ 58. We now start the presentation of the main topic of this section. First a formula will be established expressing the non-Euclidean distance between two points of the Poincaré model (see §§ 50-52).

Let u, v be two points of the upper half plane. A non-Euclidean line through u, v is depicted by the Euclidean semicircle passing through these points and orthogonal to the x -axis. Denote by s and t the points where this semicircle meets the x -axis (Fig. 76; recall that the points of the x -axis, including s and t do not belong to the Poincaré model). If the semicircle orthogonal to the x -axis and passing through u and v degenerates into a (Euclidean) line, then s denotes the point where this line meets the x -axis and t the point at infinity (Fig. 77). Consider the cross ratio $(uvst)$. It may be easily shown that $(uvst)$ is a real positive number. This will be shown first for the case of Fig. 76 (s is assumed to lie on the left of t). Denote by r_1 and r_2 the moduli of $u - s$ and $u - t$, respectively, and by θ_1 and θ_2 their arguments. Since $\angle (sut)$ is a right angle,

$$\theta_2 - \theta_1 = \pi/2$$

Accordingly,

$$\frac{u-s}{u-t} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)} = \frac{r_1}{r_2} e^{-i\pi/2}$$

and, similarly,

$$\frac{v-s}{v-t} = \frac{r_3}{r_4} e^{-i\pi/2}$$

where $r_3 = |v - s|$ and $r_4 = |v - t|$. Therefore

$$(uvst) = \frac{r_1}{r_2} \div \frac{r_3}{r_4} > 0 \quad (9)$$

In the case corresponding to Fig. 77, the numbers $u - s$ and $v - s$ are real and of the same sign; accordingly, in this case

$$(uvst) = (uvs\infty) = \frac{u-s}{v-s} = \frac{r_1}{r_3} > 0 \quad (10)$$

Thus $(uvst) > 0$ and we are justified in taking the logarithm of $(uvst)$, where the logarithm is understood in the sense of elementary algebra.

We shall now show that *the non-Euclidean distance between any two points u and v of the Poincare model is given by the formula*

$$\rho(uv) = R |\ln(uvst)| \quad (11)$$

with R a positive constant (whose choice amounts to choosing a scale).

To prove this, we shall show that $\rho(uv)$ satisfies the three conditions required by the length of a segment (see Definition 12 of § 20).

(1) Let u' and v' be a pair of points of the upper half plane determining a non-Euclidean segment that is congruent to the non-Euclidean segment determined by the points u and v . Suppose that points s' and t' bear the same relation to points u' and v' as the points s and t to points u and v . According to the definition of the congruence of non-Euclidean segments of the Poincare model (see § 52), if the non-Euclidean segment uv is congruent to the non-Euclidean segment $u'v'$, then there is a chain of inversions whose product maps points u, v, s, t into points u', v', s', t' . As shown in § 51, the product of any number of inversions is a linear-fractional transformation of either the first or the second kind; in either case the transformation is nondegenerate because each inversion maps distinct points into distinct points. In the first case, $(u'v's't') = (uvst)$ (§ 55) and in the second $(u'v's't') = (\overline{uvst})$ (§ 56). But earlier in this subsection it was shown that $(uvst)$ is a real number; hence $(\overline{uvst}) = (uvst)$. Thus in both cases, $(u'v's't') = (uvst)$. This, in view

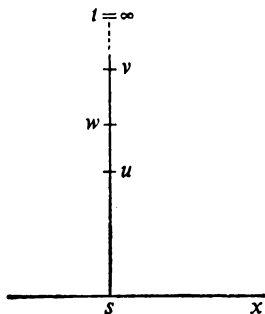


Figure 77

of formula (11), shows that $\rho(u'v') = \rho(uv)$. Note finally that $(uvst) \neq 1$ (this follows from formulas (9) and (10)); accordingly, $\ln(uvst) \neq 0$ and $\rho(uv) > 0$. Thus with every non-Euclidean segment there is associated a positive number given by formula (11). and with congruent segments there are associated equal numbers. Hence the first condition in Definition 12 of § 20 is satisfied.

(2) Let w be any point inside the non-Euclidean segment uv (Figs. 76, 77). A straight forward computation shows that

$$(uvst) = (uwst)(wvst)$$

and that $(uwst)$ and $(wvst)$ are either both greater than unity or less than unity (this can easily be shown with the aid of formulas (9) and (10)). These results imply that

$$\ln(uvst) = \ln(uwst) + \ln(wvst)$$

and in the right-hand side the logarithms are either both positive or negative. Thus

$$|\ln(uvst)| = |\ln(uwst)| + |\ln(wvst)|$$

and, according to formula (11),

$$\rho(uv) = \rho(uw) + \rho(wv)$$

which shows that the second requirement of Definition 12 of § 20 is also satisfied.

(3) If v approaches u along a non-Euclidean line, then $(uvst) \rightarrow 1$; if v approaches t , then $(uvst) \rightarrow 0$ (see Fig. 76 and formula (9), where r_1, r_2, r_3, r_4 denote the Euclidean distances between the points u and s , u and t , v and s , v and t). In the first case, $\ln(uvst) \rightarrow 0$, while in the second $\ln(uvst) \rightarrow -\infty$; accordingly, in the first case, $\rho(uv) \rightarrow 0$ and in the second $\rho(uv) \rightarrow +\infty$.

From formula (11) it is clear that $\rho(uv)$ continuously depends on v . This and the preceding results imply that $\rho(uv)$ assumes all the values between 0 and $+\infty$; in particular, there is a pair u and v for which $\rho(uv) = 1$. This shows that the third requirement in Definition 12 of § 20 is also met.

Thus we have shown that the number $\rho(uv)$ associated, according to formula (11), with any pair of points u and v is the length of a non-Euclidean segment uv (in some unit), or, in other words, is the non-Euclidean distance between points u and v .

If the measuring segment u_1v_1 is provided beforehand, then constant R in (11) must be determined subject to $\rho(u_1v_1) = 1$.

§ 59. Now we shall obtain Lobachevsky's formula expressing the function $\Pi(x)$ in terms of elementary functions of the argument x . Recall that $\alpha = \Pi(x)$ is the angle of parallelism corresponding

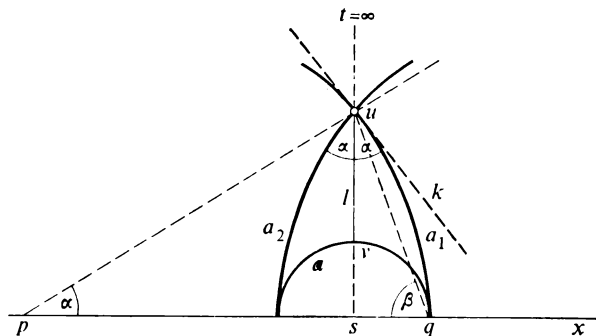


Figure 78

to a segment of length x (see § 33 and Fig. 46). Since in this section x denotes the abscissa of a point of the Poincare model, in order to avoid confusion the argument of Lobachevsky's function will be denoted by l .

The angle of parallelism corresponding to a given segment is determined by the length of the given segment and is independent of its position; therefore, in deriving the desired formula, any non-Euclidean segment of the Poincare model can be chosen so as to make the computations as simple as possible. Keeping this in mind, we shall assume the non-Euclidean segment to be represented in the Poincare model by a segment of the Euclidean line perpendicular to the x -axis. Let u and v be the endpoints of the chosen segment, s the point of intersection of line uv with the x -axis (Fig. 78), the point u is assumed to lie above v on the model and the Euclidean distance from v to the x -axis is assumed to be unity. Other necessary objects are explained in Fig. 78. The letter a denotes the semicircle representing the non-Euclidean line perpendicular to segment uv at its endpoint v , while a_1 and a_2 denote semicircles representing non-Euclidean lines passing through u and parallel to a ; p is the centre of semicircle a_1 and q the common (right) endpoint of semicircles a and a_1 ; α denotes both of the angles formed by semicircles a_1 and a_2 with segment uv . Since in the Poincare model the non-Euclidean magnitude of an angle coincides with its Euclidean magnitude, α is the angle of parallelism for the segment uv . Our aim is to express α in terms of l , where l is the non-Euclidean length of uv .

Denoting the Euclidean distance between points u and s by h , we have $u - s = h$ and $v - s = 1$, and by formula (11)

$$l = R |\ln(uvs\infty)| = R \left| \ln \frac{u-s}{v-s} \right| = R |\ln h|$$

Since $h > 1$, it follows that $\ln h > 0$; accordingly,

$$l = R \ln h \quad (12)$$

Note that $\angle upq = \alpha$ (to see this, observe that α is the angle between segment uv and semicircle a_1 , that is, the angle between segment uv and tangent uk to the semicircle a_1 ; clearly, $\angle upq = \angle suk$ because these angles are acute, and $us \perp pq$, $uk \perp up$). Next, since Δupq is isosceles, it follows that $\angle uqp = \beta = \pi - \alpha/2$. From triangle usq , we have

$$h = \tan \beta = \cot \alpha/2 \quad (13)$$

Formulas (12) and (13) yield

$$l = R \ln \cot \alpha/2$$

or

$$\alpha = 2 \arctan e^{-l/R}$$

But, since $\alpha = \Pi(l)$, we finally have

$$\Pi(l) = 2 \arctan e^{-l/R} \quad (14)$$

This is the formula of Lobachevsky the derivation of which was our aim. In Lobachevskian geometry a fundamental role is played by formula (14), since it gives the exact expression for the angle of parallelism corresponding to a segment of length l .

§ 60. In the Lobachevskian plane we shall examine segments with a length not exceeding a positive number L . Setting

$$\alpha_0 = 2 \arctan e^{-L/R}$$

we have

$$\alpha_0 \leq \Pi(l) < \pi/2$$

for $l \leq L$.

The quantity α_0 can be as close to $\pi/2$ as one likes if R is sufficiently large in comparison to L . Therefore, for all segments of length $l \leq L$, the angle of parallelism $\Pi(l)$ will be close to $\pi/2$. In other words, if a part of the Lobachevskian plane is taken where the distance between all possible points does not exceed L , the larger R is in comparison to L the less the non-Euclidean character of the Lobachevskian plane is displayed. In view of this, R may be looked upon as the "measure of non-Euclideaness" of the Lobachevskian plane; the segment of length R is called the *radius of curvature of the Lobachevskian plane*. The actual value of R , depends, of course, on the choice of scale; a proper choice of scale ensures $R = 1$. However, in itself the radius of curvature for every specific model of Lobachevskian geometry is a segment defined to within congruent displacements. To cite an example, for the Poincare model the radius of curvature is segment uv subject to $\ln(uvst) = 1$. A general

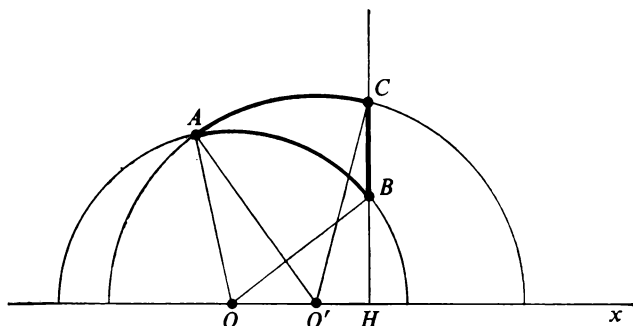


Figure 79

description of the radius of curvature, that is, a description independent of the choice of the model of Lobachevskian geometry can be found in § 216.

§ 61. Here we shall obtain fundamental relationships of Lobachevskian trigonometry assuming, as before, that the Lobachevskian plane is represented by the Poincaré model.

Let ABC be an arbitrary triangle, α, β, γ the magnitudes of angles at the vertices A, B, C , and a, b, c the non-Euclidean lengths of sides BC, AC, AB .

Applying congruent displacement, we place this triangle in a position where line BC is represented by a Euclidean half line perpendicular to the x -axis; the point where this half line meets the x -axis is called H (Fig. 79). Non-Euclidean lines AB and AC are represented by certain Euclidean semicircles with their centres on the x -axis, call these centres O and O' . Without loss of generality, we may assume that B lies between H and C . Then

$$a = R \ln (HC/HB) \quad (1)$$

where HC and HB are Euclidean lengths of segments (formula (1) may be proved in the same way as formula (12) of § 60). Further calculations are based on formula (1).

First we obtain a formula expressing the sides of a triangle in terms of its angles.* From formula (1) we have

$$\cosh \frac{a}{R} = \frac{1}{2} \left(\frac{HC}{HB} + \frac{HB}{HC} \right) = \frac{HB^2 + HC^2}{2HB \cdot HC} \quad (2)$$

Writing the obvious Euclidean formulas

$$HB^2 = OB^2 - OH^2 = OA^2 - OH^2$$

$$HC^2 = O'C^2 - O'H^2 = O'A^2 - O'H^2$$

* The derivation of the formulas of Lobachevskian geometry given below was communicated to the author for the fourth Russian edition of this book by the Vietnamese mathematician Nguyễn-Ca'nh-Toàn.

we have

$$\begin{aligned}
 HB^2 + HC^2 &= (OA^2 + O'A^2) - (OH^2 + O'H^2) \\
 &= (OO'^2 + 2OA \cdot O'A \cos (\angle OAO')) \\
 &\quad - ((OH - O'H)^2 + 2OH \cdot O'H) \\
 &= 2OB \cdot O'C \cos (\angle OAO') - 2OH \cdot O'H \quad (3)
 \end{aligned}$$

(For convenience, the computations are carried out with the aid of Fig. 79, where O' lies between O and H .) From (2) and (3) we find that

$$\cosh \frac{a}{R} = \frac{OB}{HB} \frac{O'C}{HC} \cos (\angle OAO') - \frac{OH}{HB} \frac{O'H}{HC} \quad (4)$$

Now note that

$$\begin{aligned}
 \frac{OB}{HB} &= \frac{1}{\sin (\angle BON)} = \frac{1}{\sin \beta}, & \frac{O'C}{HB} &= \frac{1}{\sin (\angle CO'H)} = \frac{1}{\sin \gamma} \\
 \angle OAO' &= \alpha, & \frac{OH}{HB} &= \cot (\angle BOH) = \cot \beta \\
 \frac{O'H}{HC} &= \cot (\angle CO'H) = \cot (\pi - \gamma) = -\cot \gamma
 \end{aligned}$$

which with the aid of (4) yield

$$\cosh \frac{a}{R} = \frac{\cos \alpha + \cos \beta \cos \gamma}{\sin \beta \sin \gamma} \quad (I)$$

Two other formulas expressing the non-Euclidean lengths of sides b and c may be obtained from (I) by a cycle permutation of the symbols α, β, γ .

Formula (I) expresses one side of a triangle in terms of its angles. The presence of such a formula means that a triangle in Lobachevskian geometry is determined by its angles; this also implies that there is no similarity of figures in Lobachevskian geometry. It is therefore natural that there is no analogue of formula (I) in Euclidean geometry.

From formula (I) it is easy to deduce for a non-Euclidean triangle the relations between sides and angles corresponding to the Euclidean sine theorem. Indeed,

$$\frac{\sinh \frac{a}{R}}{\sin \alpha} = \frac{\sqrt{\cosh^2 \frac{a}{R} - 1}}{\sin \alpha} = \frac{\sqrt{(\cos \alpha + \cos \beta \cos \gamma)^2 - \sin^2 \beta \sin^2 \gamma}}{\sin \alpha \sin \beta \sin \gamma} \quad (5)$$

Setting

$$\begin{aligned}
 Q &= (\cos \alpha + \cos \beta \cos \gamma)^2 - \sin^2 \beta \sin^2 \gamma \\
 &= \cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma + 2 \cos \alpha \cos \beta \cos \gamma - 1
 \end{aligned}$$

we find that this expression is symmetric with respect to α, β, γ ; accordingly, the right side of (5) has the same symmetry. Thus we obtain

$$\frac{\sinh \frac{a}{R}}{\sin \alpha} = \frac{\sinh \frac{b}{R}}{\sin \beta} = \frac{\sinh \frac{c}{R}}{\sin \gamma} = \frac{\sqrt{Q}}{\sin \alpha \sin \beta \sin \gamma} \quad (\text{II})$$

Using formula (I), one may also obtain formulas expressing angles of a triangle in terms of its sides. To this end, write

$$\begin{aligned} \cosh \frac{b}{R} \cosh \frac{c}{R} &= \frac{(\cos \beta + \cos \gamma \cos \alpha)(\cos \gamma + \cos \alpha \cos \beta)}{\sin^2 \alpha \sin \beta \sin \gamma} \\ \sinh \frac{b}{R} \sinh \frac{c}{R} \cos \alpha &= \frac{Q \cos \alpha}{\sin^2 \alpha \sin \beta \sin \gamma} \end{aligned}$$

whence

$$\begin{aligned} \cosh \frac{b}{R} \cosh \frac{c}{R} - \sinh \frac{b}{R} \sinh \frac{c}{R} \cos \alpha &= \frac{(1 - \cos^2 \alpha) \cos \alpha + (1 - \cos^2 \alpha) \cos \beta \cos \gamma}{\sin^2 \alpha \sin \beta \sin \gamma} \\ &= \frac{\cos \alpha + \cos \beta \cos \gamma}{\sin \beta \sin \gamma} = \cosh \frac{a}{R} \end{aligned}$$

Thus we have the following:

$$\cos \alpha = \left(\cosh \frac{b}{R} \cosh \frac{c}{R} - \cosh \frac{a}{R} \right) \left(\sinh \frac{b}{R} \sinh \frac{c}{R} \right)^{-1} \quad (\text{III})$$

If formulas (I) and (III) are compared, it will be found that there is a definite relation between the sides and angles of a triangle in Lobachevskian geometry.

Let us find the relationships between the sides and angles of a right triangle. To do this, it is enough to set, for example, $\gamma = \pi/2$ in formulas of the form (I), (II), (III). Thus we have

$$(1) \quad \sinh \frac{a}{R} = \sinh \frac{c}{R} \sin \alpha, \quad \tanh \frac{a}{R} = \tanh \frac{c}{R} \cos \beta$$

which connects a leg, the hypotenuse and one of the acute angles;

$$(2) \quad \tanh \frac{a}{R} = \sinh \frac{b}{R} \tan \alpha$$

which connects two legs and one acute angle;

$$(3) \quad \cosh \frac{c}{R} = \cosh \frac{a}{R} \cosh \frac{b}{R}$$

which connects the hypotenuse and the legs (an analogue of the Pythagorean theorem). We also note the formulas

$$\cosh \frac{c}{R} = \cot \alpha \cot \beta, \quad \cosh \frac{a}{R} = \frac{\cos \alpha}{\sin \beta}$$

expressing one side of a right angled triangle in terms of angles and hence have no analogue in Euclidean geometry.

§ 62. Rewriting formula (III), we have

$$\cosh \frac{a}{R} = \cosh \frac{b}{R} \cosh \frac{c}{R} - \sinh \frac{b}{R} \sinh \frac{c}{R} \cos \alpha \quad (A)$$

In this form formula expresses one side of any non-Euclidean triangle in terms of the other two sides and the cosine of the opposite angle.

Let us compare the last formula with the formula known from spherical trigonometry:

$$\cos \frac{a}{R} = \cos \frac{b}{R} \cos \frac{c}{R} + \sin \frac{b}{R} \sin \frac{c}{R} \cos \alpha \quad (B)$$

where R is the radius of the sphere. From this and two other similar formulas obtained by cyclic permutation of sides and angles the remaining formulas of spherical trigonometry can be obtained.

Formula (B) reduces to (A) if R is replaced by iR ($i = \sqrt{-1}$).

This feature is expressed by saying that Lobachevskian trigonometry may be considered as trigonometry on a sphere with an imaginary radius.

The deep connection of Lobachevskian geometry with spherical geometry (and also with Riemannian geometry, to which the next section is devoted) will be considered in detail in Chapter 8.

3.11. Riemannian geometry: a brief survey

§ 63. In the previous sections more than once we mentioned spherical geometry together with Euclidean and Lobachevskian geometries. Comparison of these geometries naturally arose when some elements of similarity between them was found (as in §§ 48, 62) or when they were examined from a general point of view (as in §§ 45-47). But now a very important difference must be pointed out between spherical geometry on the one hand and Lobachevskian and Euclidean geometries on the other. Namely, in both Lobachevskian and Euclidean planes two lines can have at most one common point; by contrast, in spherical geometry, where the role of lines is played by great circles of the sphere (see § 45), two "lines" (that is, two great circles) always intersect at two diametrically opposite points.* Hence, in spherical geometry, one of the most fundamental features of Euclidean and Lobachevskian geometries, namely that through two distinct points there passes only one line, fails to be valid.

There is a geometric system which to a large extent is similar to spherical geometry but where the fundamental feature of elemen-

* Such points will be called *antipodal* points.

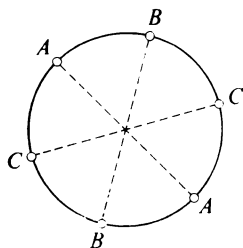


Figure 80

tary geometry is retained. This geometric system is known as *Riemannian geometry* (already mentioned in § 40). Riemannian geometry necessarily supplements Euclidean and Lobachevskian geometries. Simultaneous investigation of these three geometries was instrumental in solving the principal geometric problems of the nineteenth century (see Chapters 6-9). The essence of Riemannian geometry is given in the following subsections.

§ 64. Consider any sphere k in the Euclidean space. Let us "identify" antipodal points of this sphere, that is, each pair of antipodal points is to be treated as one object; this object will be referred to as a "point" in some special geometry, which will now be described. Every great circle of the sphere k is treated as a "line".

A "point" A is said to lie on "line" a (or "line" a is said to pass through "point" A) if the usual points of sphere k constituting "point" A lie on the great circle that represents "line" a . Then it is evident that (1) through any two "points" there passes a "line", (2) through any two "points" there passes just one "line", and (3) on every "line" there are at least two "points" (even infinitely many "points"); there are three "points" that do not lie on one "line".

Thus for this set of "points" and "lines" all the axioms of incidence for elementary plane geometry (see § 12, Axioms I.1, I.2, I.3) are satisfied. By contrast, the axioms of betweenness in the form formulated for elementary geometry are inapplicable now. The reason is that these axioms characterize the notion of the position of an ordinary point between two other ordinary points on an ordinary line. But for our "points" on our "line" the betweenness concept loses its meaning because taking any three "points" A, B, C on a "line" (that is, three pairs of antipodal points of a great circle, Fig. 80) we cannot distinguish the location of any one of them with respect to the other two.

In order to examine the betweenness relationship between the "points" of a "line", one must consider four "points". Let A, B, C, D , be four "points" of a "line"; let them be numbered in the order of their appearance (irrespective of their position on the "line"). Two-

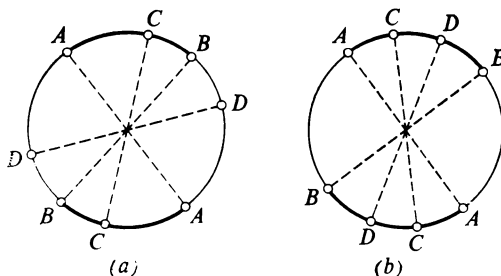


Figure 81

totally different cases are possible depending on the location of "points" A, B, C, D keeping in view their order: (a) the first two "points" A and B divide the two "points" C and D (then C and D divide A and B ; Fig. 81a), (b) the first two "points" A and B do not divide the last two "points" C and D (then C and D do not divide A and B ; Fig. 81b). Similarly, if a, b, c, d are four "lines" through one "point" and are taken in the order of their appearance, then two cases concerning their locations are possible: (i) "lines" a and b divide "lines" c and d (then c and d divide a and b ; Fig. 82a), (ii) "lines" a and b do not divide "lines" c and d (then c and d do not divide a and b , Fig. 82b). The notion of dividing "points" and "lines" will be treated as fundamental and the remaining notions concerning the location of "points" on a "line" and "line" passing through a "point" will be given in terms of it.

Let A and B be any two "points" of a "line" u . Then all the "points" of "line" u , excluding A and B , may be uniquely classified into two classes such that any two "points" of the same class do not divide A and B , while any two "points" from different classes divide A and B . In this case, we shall say that "points" A and B determine two "segments" on "line" u ; the interior points of one "segment" will be "points" belonging to one of the two classes, while the interior points of the second "segment" will be "points" belonging to the other class (in Fig. 81a and b, one of the two segments determined by "points" A and B is represented by two heavy arcs, in Fig. 81a, C is the interior "point" of this "segment" and D the interior "point" of the second "segment"; in Fig. 81b both C and D are interior "points" of one "segment").

Similar statements may be made concerning "lines" passing through one "point". Namely, if a and b are two "lines" passing through one "point" U , then all the "lines" through U , excluding a and b , may be classified into two classes such that any two "lines" of the same class do not divide a and b , while two "lines" from different classes do. In this connection, it will be said that two "lines" a and b deter-

mine two "angles" with a vertex U ; by the interior "lines" of one "angle" will be meant "lines" of one of the two classes mentioned just now, while "lines" of the second class will be treated as "lines" of the second "angle".

After this step, one may define in a natural way a triangle, the interior angles of a triangle, the interior region of a triangle, a polygon, a simple polygon (without selfintersection), the interior angles of a simple polygon and various other notions of elementary geometry.

Finally, we call two "segments" congruent if there is a motion of sphere k about itself or if there is a reflection of this sphere in a certain diametral plane such that one of these "segments" is superposed on the other (that is, the extreme and interior points of one segment fall respectively on the extreme and interior points of the other). The congruence of "angles" and that of any figures are defined analogously (a figure M treated as a set of "points" and "lines" is congruent to another figure M' if between the "points" of these figures as well as between the "lines" a one-to-one correspondence may be established such that through some motion of sphere k about itself or through a reflection in a diametral plane all the "points" and "lines" of M are superposed on corresponding "points" and "lines" of M').

Summing up, we consider (1) the connection of "points" and "lines", (2) the order of "points" on any "line" and of "lines" passing through any "point", (3) the congruence of "segments", "angles" and other figures. The system of theorems stating these relationships is known as *Riemannian geometry*; the set of "points" and "lines" understood in the aforementioned sense and satisfying the above relationships is called the *Riemannian plane*. All the theorems of Riemannian geometry are suitable interpretations of the theorems of Euclidean

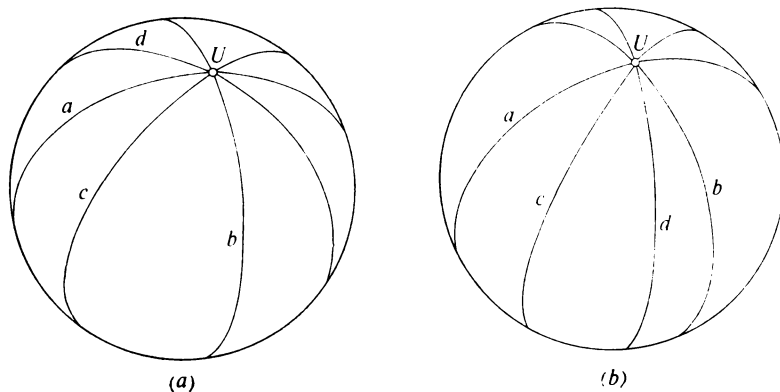


Figure 82

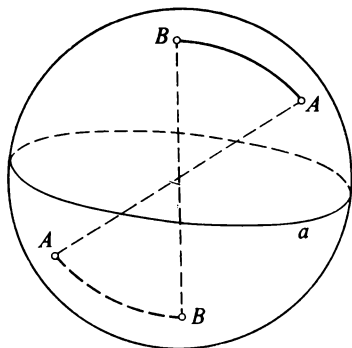


Figure 83

geometry because “points” and “lines” of the Riemannian plane are Euclidean objects.

§ 65. Let us point out a few theorems of Riemannian geometry. First of all, as noted earlier, in this geometry three axioms of connection of Euclidean planimetry are realized; in particular, any two points determine one and only one line through them. On the other hand, the following proposition is valid in Riemannian geometry but is not valid either in Euclidean or in Lobachevskian geometry: any two lines have (one) common point (this is clear, because any two great circles of the sphere have in common a pair of antipodal points). In other words, there are no parallel lines in the Riemannian plane. Whereas Euclidean geometry assumes a postulate stating that only one line passes through a given point not intersecting the given line and Lobachevskian geometry negates this postulate replacing it by the assumption that there are an infinite number of such lines, Riemannian geometry negates that postulate in another way: in this geometry any one line intersects the other.

The location of lines on the Riemannian plane differs fundamentally in another respect from that on the Euclidean or Lobachevskian planes; namely, a line does not divide the Riemannian plane into two parts. This means that whatever the line a and two points A and B lying outside this line, it is always possible to join points A and B by a segment not intersecting line a (Fig. 83).

In Riemannian geometry, the comparison of segments (among themselves) and angles (among themselves) as well as the measurement of segments and angles are carried out in a natural way (see §§ 48, 20, 21, where these notions were defined for Euclidean geometry). At the same time one may formulate and prove theorems regarding relationships between geometric quantities more or less analogous to the known theorems of Euclid and Lobachevsky.

It is interesting to compare the proposition regarding the sum of the interior angles of a triangle in Euclidean, Lobachevskian and Riemannian geometries: in Euclidean geometry the sum of the interior angles of a triangle is two right angles, in Lobachevskian geometry this sum is less than two right angles, while in Riemannian geometry the sum is greater than two right angles. In order to show the latter result, that is, that on the Riemannian plane the sum of the interior angles of a triangle is greater than two right angles, it suffices to observe that the lines of the Riemannian plane are great circles of a sphere and, since the sum of the interior angles of a spherical triangle is greater than two right angles, the same is also valid for a triangle in the Riemannian plane.

Finally, we point out that the metric relationships in Riemannian geometry are expressed by suitably interpreting the formulas of spherical geometry (this is obvious, for each figure M of the Riemannian plane is a pair of figures M_1 and M_2 of a sphere which are situated symmetrical with respect to the centre of this sphere, and each such pair of antipodal points of M_1 and M_2 is treated as one point of M ; thus every metric relationship between elements of M coincides with the metric relationship between corresponding elements of M_1 or M_2). For example, in the Riemannian plane the side a of a triangle is expressed in terms of the other two sides b and c and the opposite angle α by the formula

$$\cos \frac{a}{R} = \cos \frac{b}{R} \cos \frac{c}{R} + \sin \frac{b}{R} \sin \frac{c}{R} \cos \alpha$$

since this formula expresses the side of a triangle on a sphere with radius R (see § 62), we note that the Riemannian plane is obtained by identifying antipodal points of this sphere (with radius R). It is clear that the number R must appear in any other metric relationship concerning such a Riemannian plane. This number R (with a given scale) obviously characterizes the Riemannian plane as well as the sphere with the aid of which this plane is defined. It is also evident that the larger the number R in comparison to the size of a certain part of the Riemannian plane the less the difference in the properties of the figures lying in this part from those of Euclidean figures. Therefore R can be looked upon as the "measure of non-Euclideaness" of the Riemannian plane. The segment of length R lying in this Riemannian plane (that is, the segment understood in the sense of Riemannian geometry) is called its *radius of curvature*.

§ 66. As already noted, all theorems of Riemannian geometry are obtained by suitably interpreting theorems of Euclidean geometry. In light of this, theorems of Riemannian geometry are derived from axioms of Euclidean geometry. Of course, not all theorems of Euclidean geometry can be interpreted to give theorems of Riemannian

nian geometry; the majority of Euclidean theorems have no connection whatsoever with those objects called points and lines of the Riemannian plane. Thus in order to obtain theorems of Riemannian geometry from axioms of Euclidean geometry, one must draw some special conclusions from these axioms.

However, it is possible to build Riemannian geometry on the basis of a special system of axioms, that is, on the basis of a number of statements (concerning the notions of incidence, betweenness and congruence of objects of the Riemannian plane) from which all the remaining propositions of Riemannian geometry may be derived in a logical manner, and each conclusion leads to some theorem of this geometry.

In that case, in proving theorems of Riemannian geometry all the properties of its objects, except those mentioned in the axioms, are rendered immaterial. This axiomatic justification of Riemannian geometry turns it into an abstract geometric system. Calling any objects "points" and "lines" and expressing by the words "lies on", "divides", "are congruent to" any mutual relationships between them satisfying the requirements of axioms, we shall obtain various concrete models of abstract Riemannian geometry. Each system of objects the mutual relationships between which satisfy the axioms of Riemannian geometry may be termed a *Riemannian plane*. Thus the sphere with identified antipodal points is just one of many different Riemannian planes.

§ 67. The axioms of Riemannian geometry will not be presented here. But that various interpretations of Riemannian geometry are possible may be demonstrated by constructing a new model of this geometry. The objects of the new model will have certain relationships with already known objects of the model on a sphere. This will enable us, without taking recourse to axioms, to obtain two models realizing the same geometry. The new model will be constructed by means of the Euclidean space.

First we complete the set of elements of Euclidean space by new elements called *points at infinity*. The nature of these new elements is not important, but in introducing them we shall assume that they are associated with the original elements; namely, it is assumed that (1) with every line a there is associated one new element called the point at infinity of the given line a , (2) parallel lines have a common point at infinity, and (3) points at infinity for nonparallel lines are distinct.

The set of all points at infinity of any plane (that is, the set of points at infinity of all lines of this plane) will be considered as a new line at infinity in the given plane. The set of all points at infinity of space will be considered as a new plane at infinity. The elements at infinity satisfying these conditions are introduced in projective geometry; therefore the space complemented with the

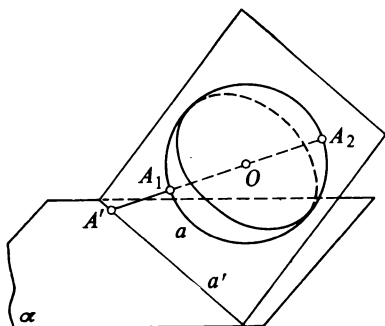


Figure 84

elements at infinity subject to the aforementioned conditions is called a *projective space*, while the plane complemented with elements at infinity subject to the same conditions is called a *projective plane* (see §§ 80-82).

Let us consider as elements of a new model the points and lines of a plane α (including its points at infinity and line at infinity). The expression "point lies on a line" is to be understood in the usual sense. Moreover, (i) all the axioms of incidence of elementary plane geometry are satisfied, (ii) any two lines intersect (possibly at the point at infinity).

Therefore the connection between points and lines of the new model satisfy the same basic conditions as the earlier model on the sphere. Now the betweenness and congruence relationships will be defined for the new model. To this end, take a sphere k and denote its centre by O (Fig. 84). Suppose that O does not lie in α . Draw any line through O ; it cuts plane α at a point A' , possibly the point at infinity, and sphere k at two antipodal points A_1 and A_2 . Treating the pair A_1 and A_2 as one point of the model of Riemannian geometry on sphere k we denote this pair by a single letter A . We shall call A' the projection of A (or A the projection of A'). Let a be some great circle of sphere k ; evidently, all pairs of its antipodal points have as their projections on plane α the points of a definite line a' (a' may be at infinity). We shall call a' the projection of a (or a the projection of a'). With each pair of antipodal points of sphere k , that is, with each point of the model of Riemannian geometry on this sphere, we associate its projection on α . With every great circle of sphere k we associate its projection on α . In other words, with every line of the model of Riemannian geometry on sphere k we associate a definite line of plane α . It is easy to see that each of these correspondences is one-to-one. It is also clear that if point A of the model of Riemannian geometry on sphere k lies on line a of

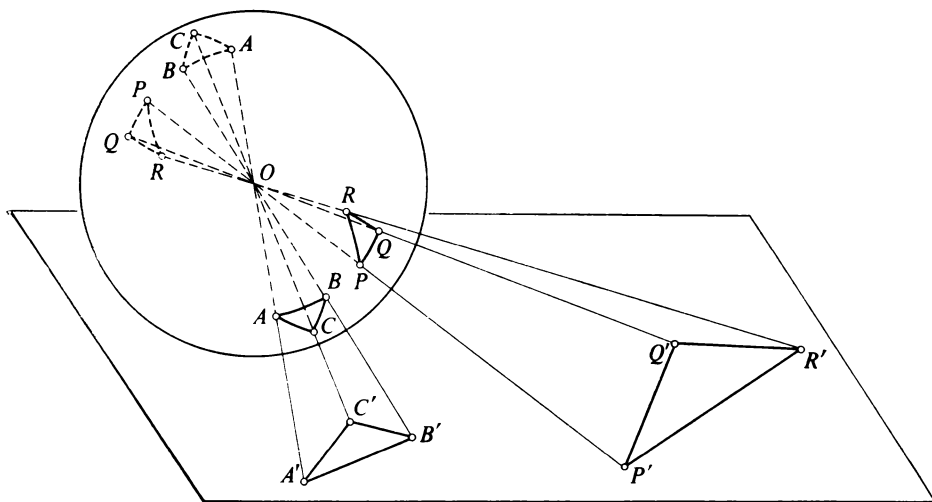


Figure 85

the same model, then in α the point A' corresponding to A lies on line a' corresponding to a .

Let A', B', C', D' be four points of plane α lying on one line u' of this plane and A, B, C, D the corresponding points of the model of Riemannian geometry on sphere k lying on one line u of this model (u and u' correspond to each other). We shall say that (a) points A' and B' divide points C' and D' on u' if A and B divide C and D on u , and (b) A' and B' do not divide C' and D' on u' if A and B do not divide C and D on u . Similarly, if a', b', c', d' are four lines of α through one point U' and a, b, c, d are the corresponding lines of the model on sphere k through point U of this model, then we shall say that (a) lines a' and b' divide lines c' and d' on plane α if a and b divide c and d , and (b) lines a' and b' do not divide lines c' and d' on α if a and b do not divide c and d . In this manner we define the "betweenness" of points on any line in α and the "betweenness" of lines of α passing through some point of this plane.

Finally, we shall say that two figures of plane α are congruent if their projections on sphere k are.

Thus for the new model we have defined the relationships of incidence, betweenness and congruence; the objects of the new model have the same mutual relationships as the corresponding objects of the model of Riemannian geometry on sphere k . This implies that every proposition regarding the incidence, betweenness and congruence for objects of the model of Riemannian geometry on sphere k will be valid for objects of the new models on the projective plane;

conversely, every proposition regarding the incidence, betweenness and congruence relationships between objects of the new model will be true for the model of Riemannian geometry on sphere k . In this manner, both models realize in a different way the same Riemannian geometry.

From the point of view of visualization, the model of Riemannian geometry on the projective plane has the advantage over the model in the form of a sphere with identified antipodal points in all cases when one speaks of connection and order between the objects because points and lines on the projective plane are represented in the way we are used to. But the model on the projective plane is worse in all cases when one speaks of congruence of figures because figures of the model on the projective plane which are congruent in the sense of Riemannian geometry are not congruent in the usual sense (see Fig. 85 showing congruent triangles ABC and PQR of the model of Riemannian geometry in the form of a sphere with identified antipodal points and corresponding, also congruent, triangles $A'B'C'$ and $P'Q'R'$ of the model of Riemannian geometry on the projective plane).

§ 68. The previous exposition concerned two-dimensional Riemannian geometry. The model of three-dimensional Riemannian geometry may be obtained by identifying antipodal points of a three-dimensional sphere in a four-dimensional Euclidean space.*

A model of three-dimensional Riemannian geometry may be constructed without taking recourse to four-dimensional space by means of projective geometry (see Chapter 6, where projective models of two-dimensional Lobachevskian and Riemannian geometries have been constructed; these constructions may be directly generalized to the three-dimensional case).

* Multidimensional Euclidean space is treated in Chapter 7; see also the first Russian edition of this book. Riemannian geometry has been discussed using the coordinate method in F. Klein: *Vorlesungen über nicht-euklidische Geometrie*, Springer, Berlin, 1963

Chapter 4

Investigation of the Axioms of Elementary Geometry

4.1. Three basic problems of axiomatic theory

§ 69. In the preceding chapter it was proved that the consistency of Euclidean geometry implies the consistency of Lobachevskian geometry. Now one may ask: How can one be certain that Euclidean geometry is consistent? Since we are treating Euclidean geometry as a system of propositions that can be derived from the axioms of Groups I-V (Chapter 2) in a logical manner, by speaking about the consistency of Euclidean geometry we mean the consistency of the system of the axioms of Groups I-V.

In the following pages it will be proved that Euclidean geometry is consistent if so is arithmetic. The question of consistency of arithmetic falls outside the scope of foundations of geometry.

In investigating the axioms of elementary geometry, we examine the following three problems:

- (1) the problem of consistency,
- (2) the problem of minimum number of axioms,
- (3) the problem of completeness.

Since these problems arise in the investigation of any system of axioms—whether it is the system of axioms of Euclidean geometry or of Lobachevskian geometry or any other system—the problems should be formulated and methods of their solutions explained in a general form.

Suppose that a system of axioms establishing mutual relationships of certain objects is given. From the accepted axioms one may draw logical conclusions regarding properties of these objects without taking into account other possible properties if they have not been mentioned in the axioms.

Thus objects of a given system of axioms may be considered objects of any nature and relationships between them mentioned in the axioms may be given any concrete meaning provided the requirements of the accepted axioms are fulfilled. Then each theorem derived in a logical way from these axioms will express a definite fact about the objects in question or, more precisely, about those properties of the objects which are mentioned in the axioms.

Any concrete choice of objects playing the role of the objects of a given system of axioms will be called the *realization* or *interpretation* of these axioms.

The set of objects itself realizing the given system of axioms will be called a "model", as was done in the previous chapter, of the logical system defined by these axioms.

If by some means axioms are realized for a model, then it is impossible to deduce by correct arguments two conclusions logically contradicting each other (say, the assertion and the negation of the same fact). Therefore, to prove the consistency of the given system of axioms it is enough to find one of its possible realizations. (If the system is inconsistent, this can be usually discovered by necessary arguments leading to a contradiction.)

The proof of the consistency of a given system of axioms may be conditional. For example, the consistency of Lobachevskian plane geometry was proved in the previous chapter by constructing Poincaré's model whose objects were chosen in the Euclidean plane. The result obtained in this way was formulated conditionally: Lobachevskian plane geometry is consistent if Euclidean plane geometry is.

The second problem is to show the necessity of all the requirements mentioned in the axioms, that is, from the accepted system of axioms it should not be possible to omit any requirement without affecting the volume of propositions on the whole (the system is minimal)*. Solving this problem completely means showing that each part of the system of axioms does not depend on those remaining, that is, any part of the system cannot be obtained from the remaining axioms in a logical way.

Denote by A one of the axioms of the (consistent) system in question. If axiom A is not a consequence of other axioms of this system, then replacing axiom A by another axiom A^* in this system stating that "assertion A is false" we must obtain a consistent system. Therefore, in order to prove that axiom A cannot be deduced from the remaining axioms of the system in question, it is enough to realize all the axioms except A on a set of objects such that in this realization axiom A does not hold.

In particular, in this manner was shown the independence of Euclid's fifth postulate from the remaining postulates of elementary geometry. Namely, for Poincaré's model, on the Euclidean half plane all the axioms of absolute geometry are realized but the Euclidean parallel axiom does not hold. Thus it is not a logical consequence of the remaining axioms. (In this case the same interpretation of geometric objects shows the independence of the Euclidean parallel postulate and the consistency of Lobachevskian geometry.)

* Since different systems of axioms may be taken as the basis for the same geometry, by omitting the superfluous conditions (if they are present) from these systems one may arrive, generally speaking, at various minimal systems. Thus a minimal system is by no means unique.

Later a similar analysis will be carried out with respect to other important axioms of elementary geometry, but, of course, the minimal problem will not be solved completely.

The formulation of the third problem of the axiomatic theory, the problem of completeness of axioms, is postponed until the end of the chapter.

§ 70. We have already seen the application of the suggested methods, in the construction of Poincaré's model. However, many minor details may have overshadowed the main points which should be made clear by taking as simple an example as possible. We shall now construct a model for Hilbert's axioms of the first group only, treating this group as a separate axiomatic system.

Consider a tetrahedron and call its vertices "points", its edges "lines" and its faces "planes". Thus the set of geometric elements in this realization is composed of only four points, six lines and four planes.

The lines and planes have a certain association with points, and if, for example, line a is associated with point A , then we say that " a passes through A " or " A lies on a ", and so on. In our realization, as in any concrete realization, these associations must be precisely described. Let us agree to call those lines and planes as passing through a point, represented by a vertex, which are represented by edges and faces that contain the given vertex.

It may be easily seen that all the requirements of Axioms I.1-I.8 are then satisfied. Let us take a look at them in a sequence.

Axiom I.1. Given any two points A and B , there is a line a passing through both A and B . This axiom is satisfied for any two vertices of the tetrahedron are joined by an edge.

Axiom I.2. Given any two points A and B , there is at most one line through both A and B . The requirement of this axiom is met as two vertices of a tetrahedron are joined by only one edge.

Axiom I.3. On every line there are at least two points; there are at least three points not lying on one line. The conditions of this axiom are satisfied since on each edge there exist two vertices and there are three vertices (even four!) not lying on one edge.

Axiom I.4. Given any three noncollinear points A , B , C , there is a plane α passing through each of the three points A , B , C ; every plane contains at least one point. Conditions of this axiom are met because through every three vertices there passes a face and every face contains a vertex (even three vertices!).

Axiom I.5. Given any three noncollinear points A , B , C , there is at most one plane passing through each of A , B , C . Our model satisfies this condition since through every three vertices only one face passes.

Axiom I.6. If two points A and B of a line a lie on a plane α , then every point of a lies on α . This axiom is satisfied. Indeed, if

two vertices of one edge lie on some face, then every vertex of this edge lies on the same face since one edge contains just two vertices.

Axiom I.7. If two planes α and β have one point A in common, then they have at least one more point B in common. The requirement of this axiom is met because any two faces have two common vertices.

Axiom I.8. There are at least four points not lying in one plane. The conditions of this axiom is also met because four vertices of the tetrahedron do not all lie on one face.

Thus, our realization satisfies the requirements of all the axioms of the first group. By the way, observe that this realization of Axioms I.1-I.8 is minimal among all possible realizations in the sense that every line has only a pair of points; the points are four in number, etc. Just as many elements are required by the axioms. True, Axiom I.4 requires every plane to have at least one point whereas in our realization every plane contains three points; however, Theorem 3 of § 12 shows that this number is the minimum.

Since a concrete realization has been pointed out for Axioms I.1-I.8, it may be said that the axioms of the first group form a consistent system.

In the preceding article, the general principle for establishing the independence of two propositions from each other was described. Here a simple illustration of this principle may be given. Let us pose, for example, the question: Is it possible to show with the aid of Axioms I.1-I.8 that the set of elements of geometry is infinite? The answer to this question is negative, since Axioms I.1-I.8 have been realized on a finite set of objects. In other words, the proposition that the set of elements of geometry is infinite does not depend on the axioms of the first group.

4.2. Consistency of the axioms of Euclidean geometry

§ 71. We shall now prove the consistency of all the five groups of axioms of Euclidean geometry.

We are used to imagining these axioms realized on a set of objects which we visualize and in our imagination they appear as abstractions of objects from the real world as observed by us. But points, lines and planes as images of our geometrical imagination cannot be described mathematically. Therefore in proving the consistency of the axioms of Euclidean geometry a model must be found which, apart from our geometric visualization of it, is meaningful. To this end, a realization of Axioms I-V will be constructed, which will be called arithmetic because its objects will be combinations of numbers. This will establish the consistency of Euclidean geometry provided arithmetic is consistent.

In order not to overshadow the main points by minor details of a computational nature, we confine ourselves to the case of planimetry, that is, only Axioms I.1-I.3 and the Axioms of Groups II-V will be taken into account.

In the arithmetical realization a "point" will mean any pair of real numbers (x, y) and a "line" the ratio of three real numbers $(u \div v \div w)$ with the condition that at least one of the two numbers u or v is not zero.*

We shall say that "a point (x, y) lies on the line $(u \div v \div w)$ " or "the line $(u \div v \div w)$ passes through the point (x, y) " if

$$ux + vy + w = 0$$

Then it is easy to check that the conditions of Axioms I.1-I.3 are satisfied.

Indeed, if (x_1, y_1) and (x_2, y_2) are two distinct points, then the ratio of three numbers $u = y_1 - y_2$, $v = x_2 - x_1$, $w = x_1y_2 - x_2y_1$ is a line (the numbers $y_1 - y_2$ and $x_2 - x_1$ cannot be both zero simultaneously since the points (x_1, y_1) and (x_2, y_2) are distinct) passing through both the points (x_1, y_1) and (x_2, y_2) because

$$ux_1 + vy_1 + w = (y_1 - y_2)x_1 + (x_2 - x_1)y_1 + (x_1y_2 - x_2y_1) = 0$$

$$ux_2 + vy_2 + w = (y_1 - y_2)x_2 + (x_2 - x_1)y_2 + (x_1y_2 - x_2y_1) = 0$$

This shows that Axiom I.1 is satisfied.

Next, the equations

$$ux_1 + vy_1 + w = 0$$

$$ux_2 + vy_2 + w = 0$$

imply that

$$u \div v \div w = (y_1 - y_2) \div (x_2 - x_1) \div (x_1y_2 - x_2y_1)$$

Thus the points (x_1, y_1) and (x_2, y_2) determine only one line $(u \div v \div w)$, and the requirements of Axiom I.2 is fulfilled.

The conditions of Axiom I.3 are also met. In fact, since the equation

$$ux + vy + w = 0$$

always has an infinite number of distinct solutions, there are not only two but infinitely many points on a line. As an example of three points not lying on one line, one may cite the points $(0, 0)$, $(1, 0)$ and $(0, 1)$; there is no line joining these points since, evidently, there are no three numbers u, v, w which are not all zero simulta-

* The ratio of three numbers u, v, w is the aggregate of numbers u, v, w with the condition that the aggregates u, v, w and $\lambda u, \lambda v, \lambda w$, where λ is any number $\neq 0$, are identical.

neously and still satisfy the equations

$$u \cdot 0 + v \cdot 0 + w = 0$$

$$u \cdot 1 + v \cdot 0 + w = 0$$

$$u \cdot 0 + v \cdot 1 + w = 0$$

Now we shall define the "betweenness" relationship. Let a line $(u \div v \div w)$ and three points (x_1, y_1) , (x_2, y_2) , (x_3, y_3) on it be given. Suppose first that $v \neq 0$. The point (x_2, y_2) will be said to lie between the points (x_1, y_1) and (x_3, y_3) if

$$x_1 < x_2 < x_3 \quad \text{or} \quad x_1 > x_2 > x_3$$

But if $v = 0$, then necessarily $x_1 = x_2 = x_3 = -w/u$ for the points of this line and the preceding conditions are inapplicable. In this case (x_2, y_2) is said to lie between (x_1, y_1) and (x_3, y_3) if

$$y_1 < y_2 < y_3 \quad \text{or} \quad y_1 > y_2 > y_3$$

The "betweenness" relationship thus defined satisfies the conditions of all the axioms of betweenness, Axioms II.1-II.4.

It is obvious that the requirements of linear Axioms II.1-II.3 are fulfilled. We shall show that the requirements of Axiom II.4 is also met.

First note the following fact: if (x_1, y_1) and (x_2, y_2) are the endpoints of a segment, then all the interior points of this segment are of the form $(\lambda x_1 + (1 - \lambda) x_2, \lambda y_1 + (1 - \lambda) y_2)$, where λ is any number satisfying $0 < \lambda < 1$. Further, if a line $(u \div v \div w)$ passes through a point of a segment having the endpoints (x_1, y_1) and (x_2, y_2) , then the numbers $ux_1 + vy_1 + w$ and $ux_2 + vy_2 + w$ have opposite signs. Indeed, if the interior point of the segment through which the line passes corresponds to λ , then

$$u [\lambda x_1 + (1 - \lambda) x_2] + v [\lambda y_1 + (1 - \lambda) y_2] + w = 0$$

whence

$$\lambda (ux_1 + vy_1 + w) = - (1 - \lambda) (ux_2 + vy_2 + w)$$

and since λ and $1 - \lambda$ are positive, the numbers $ux_1 + vy_1 + w$ and $ux_2 + vy_2 + w$ have opposite signs.

Now let $A(x_1, y_1)$, $B(x_2, y_2)$ and $C(x_3, y_3)$ be three points not lying on one line and $(u \div v \div w)$ be the line passing through none of these points. We must show that if the line $(u \div v \div w)$ passes through a point of the segment AB , then it passes through either a point of the segment AC or a point of the segment BC .

Since the line $(u \div v \div w)$ does not pass through any of the points A, B, C , the numbers

$$\alpha_1 = ux_1 + vy_1 + w, \quad \alpha_2 = ux_2 + vy_2 + w, \quad \alpha_3 = ux_3 + vy_3 + w$$

are nonzero, and, by the preceding result, α_1 and α_2 have opposite signs. Suppose that the sign of α_3 is different from that of α_1 ; then the line $(u \div v \div w)$ intersects the segment AC . To establish this, take a number λ determined by the equation $\lambda\alpha_1 + (1 - \lambda)\alpha_3 = 0$, that is,

$$\lambda = \alpha_3 \div (\alpha_3 - \alpha_1)$$

Noting that α_1 and α_3 have opposite signs, we find that

$$\lambda = \frac{\alpha_3}{\alpha_3 - \alpha_1} = \frac{|\alpha_3|}{|\alpha_3| + |\alpha_1|}$$

whence $0 < \lambda < 1$. Consequently, the point (x, y) , where

$$x = \lambda x_1 + (1 - \lambda)x_3, \quad y = \lambda y_1 + (1 - \lambda)y_3$$

is a point of the segment AC . On the other hand, the point (x, y) lies on the line $(u \div v \div w)$ because

$$\begin{aligned} ux + vy + w &= \lambda(ux_1 + vy_1 + w) + (1 - \lambda)(ux_3 + vy_3 + w) \\ &= \lambda\alpha_1 + (1 - \lambda)\alpha_3 = 0 \end{aligned}$$

Thus the line $(u \div v \div w)$ indeed intersects the segment AC . Similarly, when α_3 and α_2 have opposite signs, it may be established that the line $(u \div v \div w)$ intersects the segment BC . But since α_1 and α_2 are of opposite signs, the sign of α_3 is necessarily opposite to that of either α_1 or α_2 . Thus the desired assertion is proved.

Now we shall define the notion of congruence. To this end, we consider a class of transformations known in algebra as orthogonal transformations. Let the following equations be given:

$$\left. \begin{aligned} x' &= a_1x + b_1y + c_1 \\ y' &= a_2x + b_2y + c_2 \end{aligned} \right\} \quad (*)$$

which, given $a_1, \dots, c_2$, map every point (x, y) into a definite point (x', y') . The transformation is called an *orthogonal transformation* if the coefficients a_1, b_1, a_2, b_2 satisfy the condition

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (**)$$

First, let us mention some properties of orthogonal transformation (*). The condition (**) yields

$$\left. \begin{aligned} a_1^2 + b_1^2 &= 1 \\ a_2^2 + b_2^2 &= 1 \\ a_1a_2 + b_1b_2 &= 0 \end{aligned} \right\} \quad (1)$$

These three equations are equivalent to equation (**) and therefore characterize the orthogonality of transformation (*).

Equations (1) first of all imply that both a_1 and a_2 as well as both b_1 and b_2 cannot be zero simultaneously. Indeed, if, for example, $a_1 = a_2 = 0$, then the third equation in (1) implies $b_1 b_2 = 0$, which together with the assumption $a_1 = a_2 = 0$ contradicts one of the first two equations of (1). Next, the equation $a_1 a_2 + b_1 b_2 = 0$ gives $a_1^2 a_2^2 = b_1^2 b_2^2$. Noting this and multiplying the first equation in (1) by b_2^2 and the second by a_1^2 and subtracting, we obtain

$$0 = b_2^2 - a_1^2$$

whence $b_2 = \delta_1 a_1$, where $\delta_1^2 = 1$. Similarly, $b_1 = \delta_2 a_2$, where $\delta_2^2 = 1$. But $b_1 b_2 = -a_1 a_2$ and hence $\delta_1 \delta_2 = -1$; thus either

$$b_1 = -a_2, \quad b_2 = a_1$$

or

$$b_1 = a_2, \quad b_2 = -a_1$$

We see that transformation (*) may be written in one of the following forms:

$$\left. \begin{aligned} x' &= \alpha x - \beta y + c_1 \\ y' &= \beta x + \alpha y + c_2 \end{aligned} \right\} \quad (\text{I})$$

or

$$\left. \begin{aligned} x' &= \alpha x + \beta y + c_1 \\ y' &= \beta x - \alpha y + c_2 \end{aligned} \right\} \quad (\text{II})$$

where α and β denote a_1 and a_2 ; in either case the orthogonality conditions (1) reduce to the relation

$$\alpha^2 + \beta^2 = 1$$

The transformations (I) and (II) are referred to respectively as orthogonal transformations of the *first* and *second kinds*.

The following property of orthogonal transformations will play an important role below: the points lying on a half line are mapped by any orthogonal transformation into points also lying on a half line. Before proving this property, we describe a convenient way of defining half lines.

Let a line a ($u \div v \div w$) and a point O (x_0, y_0) on it be given; since O lies on line a , we have

$$ux_0 + vy_0 + w = 0$$

If M (x, y) is any point of the line a , then similarly

$$ux + vy + w = 0$$

Whence

$$u(x - x_0) + v(y - y_0) = 0$$

Setting $m = \lambda v$, $n = -\lambda u$, where λ is any number $\neq 0$, we can write the last equation in the form

$$\frac{x-x_0}{m} = \frac{y-y_0}{n}$$

Denoting each of the equal ratios by t , we find

$$\left. \begin{aligned} x &= x_0 + mt \\ y &= y_0 + nt \end{aligned} \right\} \quad (2)$$

For each value of t equations (2) determine a certain point of the line, moreover, to various numerical values of t of the same sign there correspond points lying on the same side of the point $O(x_0, y_0)$, while to numerical values of t with opposite signs there correspond points lying on opposite sides of O . This becomes immediately clear if one examines the definition of "betweenness" given above.

Thus to positive values of t there correspond points of one of the two half lines into which O divides the line a ; to negative values of t there corresponds points of the second half line.

A more convenient way of defining the points of a half line is to use equation (2) in which t is always positive, and then to distinguish between the half lines with a common vertex at O of line a according to the signs of the quantities m and n ; if for one of the half lines $m = p$ and $n = q$, then for the other $m = -p$, $n = -q$.

The quantities m and n are called *normalized parameters* of a half line if they satisfy

$$m^2 + n^2 = 1$$

When $\lambda = \pm (u^2 + v^2)^{-1/2}$, this equation is satisfied.

A half line is clearly determined by the vertex (x_0, y_0) and the normalized parameters m and n . Conversely, as soon as a half line is given, its vertex (x_0, y_0) and the normalized parameters m and n are uniquely determined. A half line will be denoted by $(x_0, y_0; m, n)$, necessarily assuming that $m^2 + n^2 = 1$.

Now it is a simple matter to establish the aforementioned property of orthogonal transformations: any orthogonal transformation maps points constituting a half line into points also constituting a half line.

Let a half line $(x_0, y_0; m, n)$ be given. All the points of this half line are obtained from the formulas

$$\begin{aligned} x &= x_0 + mt \\ y &= y_0 + nt \end{aligned}$$

in which t assumes all possible positive values. Consider an orthogonal transformation of the first kind

$$\begin{aligned} x' &= \alpha x - \beta y + c_1 \\ y' &= \beta x + \alpha y + c_2 \end{aligned}$$

or of the second kind

$$x' = \alpha x + \beta y + c_1$$

$$y' = \beta x - \alpha y + c_2$$

Any point (x, y) of the given half line is mapped into a point

$$x' = (\alpha m - \beta n) t + (\alpha x_0 - \beta y_0 + c_1)$$

$$y' = (\beta m + \alpha n) t + (\beta x_0 + \alpha y_0 + c_2)$$

in the first case, and into a point

$$x' = (\alpha m + \beta n) t + (\alpha x_0 + \beta y_0 + c_1)$$

$$y' = (\beta m - \alpha n) t + (\beta x_0 - \alpha y_0 + c_2)$$

in the second. In either case these expressions may be put in the form

$$x' = m' t + x'_0$$

$$y' = n' t + y'_0$$

and, accordingly, points (x', y') obtained by varying t lie on the half line with parameters m', n' . This proves the desired property. Note that the parameters

$$m' = \alpha m - \beta n$$

$$n' = \beta m + \alpha n$$

or

$$m' = \alpha m + \beta n$$

$$n' = \beta m - \alpha n$$

are normalized. Indeed,

$$\begin{aligned} m'^2 + n'^2 &= (\alpha m \mp \beta n)^2 + (\beta m \pm \alpha n)^2 \\ &= (\alpha^2 + \beta^2) m^2 + (\alpha^2 + \beta^2) n^2 = m^2 + n^2 = 1 \end{aligned}$$

We shall say that the half line $(x'_0, y'_0; m', n')$ is obtained from the half line $(x_0, y_0; m, n)$ by an orthogonal transformation. Thus we have the following assertion:

An orthogonal transformation of the first kind (I) or the second kind (II) of points (x, y) into points (x', y') is determined by the orthogonal transformation of the first kind or the second kind, respectively, of half lines $(x_0, y_0; m, n)$ into half lines $(x'_0, y'_0; m', n')$.

The quantities $x'_0, y'_0; m', n'$ are expressed by the formulas

$$\left. \begin{aligned} x'_0 &= \alpha x_0 - \beta y_0 + c_1 \\ y'_0 &= \beta x_0 + \alpha y_0 + c_2 \\ m' &= \alpha m - \beta n \\ n' &= \beta m + \alpha n \end{aligned} \right\} \quad (\text{I}^*)$$

in the case of a transformation of the first kind and by

$$\left. \begin{aligned} x'_0 &= \alpha x_0 + \beta y_0 + c_1 \\ y'_0 &= \beta x_0 - \alpha y_0 + c_2 \\ m' &= \alpha m + \beta n \\ n' &= \beta m - \alpha n \end{aligned} \right\} \quad (\text{II}^*)$$

in the case of a transformation of the second kind; moreover, if the points (x, y) lie on the half line $(x_0, y_0; m, n)$ their images (x', y') lie on the half line $(x'_0, y'_0; m', n')$.

Now we are in a position to define congruence of segments and angles for our realization.

A segment AB will be said to be congruent to the segment $A'B'$ if there is an orthogonal transformation (of points) mapping A into A' and B into B' .

An angle (h, k) will be said to be congruent to the angle (h', k') if there is an orthogonal transformation (of half lines) mapping the half line h into the half line h' and the half line k into the half line k' .

We have to show that these definitions obey Axioms III.1-III.5. To this end, we examine successively all the axioms of the third group.

Axiom III.1 requires that for any preassigned segment AB there is just one point B' on a given side of a preassigned point A' of a line a' determining segment $A'B'$ such that AB is congruent to $A'B'$.

Let the segment $A(x_0, y_0)B(x, y)$ and a point $A'(x'_0, y'_0)$ on a line a' ($u' \div v' \div w'$) be given. The quantities

$$m' = \frac{v'}{\sqrt{u'^2 + v'^2}}, \quad n' = \frac{-u'}{\sqrt{u'^2 + v'^2}}$$

are normalized parameters of one of the two half lines determined on line a' by the point A' (the quantities $-m'$ and $-n'$ will be normalized parameters for the second half line).

If we let m and n denote the normalized parameters of the half line AB , then the point $B(x, y)$ is determined by formulas

$$x = x_0 + mt$$

$$y = y_0 + nt$$

for some definite positive t .

We shall find an orthogonal transformation mapping the half line $(x_0, y_0; m, n)$ into the half line $(x'_0, y'_0; m', n')$. Using (I*), we see that the equations

$$\alpha m - \beta n = m'$$

$$\beta m + \alpha n = n'$$

immediately yield

$$\alpha = mm' + nn'$$

$$\beta = mn' - nm'$$

What is more,

$$\begin{aligned}\alpha^2 + \beta^2 &= (mm' + nn')^2 + (mn' - nm')^2 \\ &= m^2(m'^2 + n'^2) + n^2(m'^2 + n'^2) \\ &= m^2 + n^2 = 1\end{aligned}$$

Determining c_1 and c_2 from the first pair of equations in (I*), we obtain just one orthogonal transformation of the first kind,

$$\begin{aligned}x' &= \alpha x - \beta y + c_1 \\ y' &= \beta x + \alpha y + c_2\end{aligned}$$

which maps half line $(x_0, y_0; m, n)$ into half line $(x'_0, y'_0; m', n')$.

It may be shown in the same way that there is exactly one transformation of the second kind which also maps the half line $(x_0, y_0; m, n)$ into the half line $(x'_0, y'_0; m', n')$.

Both the transformations map the point $B(x, y)$ into the same point $B'(x', y')$:

$$\begin{aligned}x' &= x'_0 + m't \\ y' &= y'_0 + n't\end{aligned}$$

Therefore there is only one point B' of line a' on a given side of A' such that $AB \equiv A'B'$, proving the truth of Axiom III.1.

Axiom III.1 further requires that

$$AB \equiv BA$$

which is also true. In fact, the orthogonal transformation

$$\begin{aligned}x' &= -x + (x_1 + x_2) \\ y' &= -y + (y_1 + y_2)\end{aligned}$$

maps the point $A(x_1, y_1)$ into the point $B(x_2, y_2)$ and, conversely, the point $B(x_2, y_2)$ into the point $A(x_1, y_1)$. Thus all the requirements of Axiom III.1 are met.

Consider now Axiom III.2, according to which the congruence relations

$$A'B' \equiv AB \quad \text{and} \quad A''B'' \equiv AB$$

imply

$$A'B' \equiv A''B''$$

For our realization this condition is met in view of the group properties of orthogonal transformations. Namely: (a) every orthogonal transformation has an inverse which is also an orthogonal

transformation, (b) if an orthogonal transformation maps points (x, y) into points (x', y') and another orthogonal transformation maps (x', y') into (x'', y'') , then the resulting transformation (that is, their product) which maps (x, y) into (x'', y'') is also orthogonal.

Indeed, consider any orthogonal transformation whose matrix is Φ . Denoting its transpose by Φ' and the identity matrix by I , the orthogonality condition (**) (see p. 194) may be written in the form

$$\Phi\Phi' = I \quad (N)$$

This implies that the value of the determinant of Φ is ± 1 and, since it is not zero, each orthogonal transformation has an inverse. The matrix of the inverse transformation satisfies the orthogonality condition; to see this, first observe that (N) implies

$$\Phi^{-1} = \Phi'$$

but $\Phi(\Phi'\Phi) = (\Phi\Phi')\Phi = \Phi$ therefore $\Phi'\Phi = I$ and

$$(\Phi^{-1})(\Phi^{-1})' = I$$

Thus the inverse transformation of an orthogonal transformation is orthogonal.

Further, if Φ and Ψ are matrices of two orthogonal transformations, then the product of these transformations has a matrix $X = \Psi\Phi$. Using the known fact

$$(\Psi\Phi)' = \Phi'\Psi'$$

it may be easily shown that X satisfies the orthogonality condition; in fact,

$$XX' = \Psi\Phi(\Psi\Phi)' = \Psi\Phi(\Phi'\Psi') = \Psi(\Phi\Phi')\Psi' = \Psi I\Psi' = \Psi\Psi' = I$$

Thus successive applications of orthogonal transformations result in an orthogonal transformation.

Having established that orthogonal transformations have group properties, it may now be easily shown that Axiom III.2 is satisfied.

Let $A'B' \equiv AB$ and $A''B'' \equiv AB$. If an orthogonal transformation maps a point M' into M , we shall write

$$M = \Phi(M')$$

If $A'B' \equiv AB$, there is a transformation $M = \Phi(M')$ such that

$$A = \Phi(A'), \quad B = \Phi(B')$$

Similarly, if $A''B'' \equiv AB$, there is a transformation $M = \Psi(M'')$ such that

$$A = \Psi(A''), \quad B = \Psi(B'')$$

Denoting by Ψ^{-1} the inverse of transformation Ψ , we have

$$A'' = \Psi^{-1}(A) = \Psi^{-1}(\Phi(A'))$$

$$B'' = \Psi^{-1}(B) = \Psi^{-1}(\Phi(B'))$$

In view of the group properties, the transformation $\Psi^{-1}\Phi$ is orthogonal; hence $A'B' \equiv A''B''$.

Let us examine Axiom III.3. Let A, B, C be three points of any line a , where B is between A and C ; let A', B', C' be three points of another line a' situated analogously. Axiom III.3 requires that

$$AB \equiv A'B', \quad BC \equiv B'C'$$

imply

$$AC \equiv A'C'$$

In the light of the arguments used in connection with Axiom III.1, there is an orthogonal transformation mapping the half line BA into the half line $B'A'$ and simultaneously the half line BC into the half line $B'C'$. Since $AB \equiv A'B'$ and $BC \equiv B'C'$, it follows from the same arguments (or from Axiom III.1 itself) that the aforementioned transformations maps the point A into A' and the point C into C' . Hence $AC \equiv A'C'$, and Axiom III.3 is fulfilled.

Now we shall show that Axiom III.4 holds in the arithmetical realization: if $\angle(h, k)$ is any given angle and h' any given half line, then there is exactly one half line k' on a given side of the half line h' such that $\angle(h, k)$ is congruent to $\angle(h', k')$; furthermore,

$$\angle(h, k) \equiv \angle(h, k), \quad \angle(h, k) \equiv \angle(k, h)$$

Here we need to clearly distinguish between orthogonal transformations of the first and the second kind.

Take a half line h and complete it to form a line $\bar{h}$, and consider the two half planes separated by line $\bar{h}$. Denote one of them by I and the other by II . Similarly, let h' be any half line contained in a line $\bar{h}'$ and let I', II' be two half planes separated by $\bar{h}'$.

Suppose that Φ_1 and Φ_2 are two orthogonal transformations of the first and second kind, respectively, each of which maps the half line h into the half line h' . Then the following result holds: each of the transformations Φ_1 and Φ_2 maps points of the half plane I into points of one of the half planes I' or II' and points of the half plane II into points of the second of the half planes I' or II' ; moreover, if Φ_1 maps half plane I into half plane I' , then Φ_2 maps I into II' .

In order to prove this result, first note that to points (x_1, y_1) and (x_2, y_2) located on opposite sides of the line $(u \div v \div w)$ there correspond numbers $ux_1 + vy_1 + w$ and $ux_2 + vy_2 + w$ with op-

posite signs (this was shown earlier while investigating Pasch's axiom). Thus for points (x, y) of one half plane $ux + vy + w > 0$ while for points of the other half plane $ux + vy + w < 0$.

If (x_0, y_0) is the vertex of the given half line, and m and n its normalized parameters, then the conditions that (x, y) belongs to one or the other half plane separated by line $\bar{h}$ may be expressed in the form

$$n(x - x_0) - m(y - y_0) > 0$$

or

$$n(x - x_0) - m(y - y_0) < 0$$

Let (x'_0, y'_0) be the vertex and m' and n' the normalized parameters for half line h' . If

$$x' = \alpha x - \beta y + c_1$$

$$y' = \beta x + \alpha y + c_2$$

is an orthogonal transformation of the first kind which maps h into h' , then

$$x' - x'_0 = \alpha(x - x_0) - \beta(y - y_0)$$

$$y' - y'_0 = \beta(x - x_0) + \alpha(y - y_0)$$

$$m' = \alpha m - \beta n$$

$$n' = \beta m + \alpha n$$

whence

$$n'(x' - x'_0) - m'(y' - y'_0) = n(x - x_0) - m(y - y_0) \quad (\alpha)$$

But if h is mapped into h' by a transformation of the second kind

$$x' = \alpha x + \beta y + c_1$$

$$y' = \beta x - \alpha y + c_2$$

then

$$n'(x' - x'_0) - m'(y' - y'_0) = -n(x - x_0) + m(y - y_0) \quad (\beta)$$

The equations (α) and (β) immediately imply the aforementioned property of orthogonal transformations; by the same token the first condition in Axiom III.4 is also verified for the realization in question.

Indeed, as already noted, there is one orthogonal transformation of the first kind and one of the second kind mapping side h of angle $\angle(h, k)$ into half line h' . Of these two transformations only one maps half line k into half line k' contained in the preassigned half plane separated by line $\bar{h}'$. Thus, on a given side of line $\bar{h}'$ there is exactly one half line k' defining $\angle(h', k')$ such that $\angle(h, k) \equiv \angle(h', k')$.

The remaining two requirements of Axiom III.4 are verified even more easily.

The relationship $\angle(h, k) \equiv \angle(k, h)$ is true because there is an orthogonal transformation under which h and k retain their positions. Such a transformation is the identity transformation

$$x' = x$$

$$y' = y$$

The relationship $\angle(h, k) \equiv \angle(k, h)$ is valid since there is an orthogonal transformation which maps h into k and k into h . Namely, if (x_0, y_0) is the vertex of an angle, and m_1, n_1 and m_2, n_2 are normalized parameters of half lines h and k , then such a transformation (of the second kind) is

$$\begin{aligned} x' &= (m_1 m_2 - n_1 n_2) x + (n_1 m_2 + m_1 n_2) y \\ &\quad + [x_0 - (m_1 m_2 - n_1 n_2) x_0 - (n_1 m_2 + m_1 n_2) y_0] \\ y' &= (n_1 m_2 + m_1 n_2) x - (m_1 m_2 - n_1 n_2) y \\ &\quad + [y_0 - (n_1 m_2 + m_1 n_2) x_0 + (m_1 m_2 - n_1 n_2) y_0] \end{aligned}$$

These formulas indeed yield $x'_0 = x_0$ and $y'_0 = y_0$, and according to formulas (II*) with given α and β we have $m'_1 = m_2$, $n'_1 = n_2$ and $m'_2 = m_1$, $n'_2 = n_1$.

This completes the verification of all the conditions in Axiom III.4.

Finally, consider Axiom III.5: if $AB \equiv A'B'$, $AC \equiv A'C'$ and $\angle BAC \equiv \angle B'A'C'$ for two triangles ABC and $A'B'C'$, then $\angle ABC \equiv \angle A'B'C'$ and $\angle ACB \equiv \angle A'C'B'$.

These conditions are satisfied for our realization. In fact, in view of the above discussion, if $AB \equiv A'B'$, there are two orthogonal transformations (one of the first kind and the other of the second) mapping point A into point A' and point B into point B' . Since $\angle BAC \equiv \angle B'A'C'$ one of these transformations maps the half line AC into the half line $A'C'$, and, because $AC \equiv A'C'$, the same transformation maps point C into point C' . Accordingly, there is an orthogonal transformation which maps the points A, B, C into points A', B', C' , respectively; hence $\angle ABC \equiv \angle A'B'C'$ and $\angle ACB \equiv \angle A'C'B'$.

Thus the proposed definitions of the congruence of segments and angles satisfy all the axioms of the third group.

We now examine axioms of continuity, Axioms IV.1-IV.2. In our list of axioms, the fourth group is composed of the axioms of Archimedes and Cantor. We could have directly verified the validity of these axioms as we did with axioms of Groups I, II, III; however, it is more profitable to act in a different way. We apply Theorem 41 of § 23, which establishes the equivalence (when the axioms of Groups I-III are satisfied) of Axioms IV.1 and IV.2 to Dedekind's

principle. In view of this theorem, it suffices to show that in the arithmetical realization Dedekind's principle is valid for the point set of every line. This may be done very easily. In fact, let $(u \div v \div w)$ be any line with, for example, $v \neq 0$; it will be assumed that a point (x_1, y_1) precedes a point (x_2, y_2) on this line if $x_1 < x_2$. Then performing a Dedekind's cut of the set of points (x, y) by the line $(u \div v \div w)$, a cut of the set of real numbers $\{x\}$ is also simultaneously performed. Since Dedekind's principle holds for the set of real numbers, there is a number $\bar{x}$ performing this cut, that is, this number closes one of the classes. Setting

$$\bar{y} = \frac{-u\bar{x} - w}{v}$$

we see that point $(\bar{x}, \bar{y})$ lies on the line $(u \div v \div w)$ and closes one of the classes of Dedekind's cut on this line. Therefore, for every Dedekind's cut of the point set of any line there is a point performing this cut. In other words, Dedekind's principle holds for any line. Then Theorem 41 of § 23 implies that the axioms of continuity, Axioms IV.1 and IV.2, hold for the arithmetical realization.

It remains to consider Axiom V, the parallel axiom.

Let $(u \div v \div w)$ be any line and (x_0, y_0) a point outside this line, that is, satisfying the condition

$$ux_0 + vy_0 + w \neq 0$$

We must find if there is only one line passing through (x_0, y_0) and not having common points with the line $(u \div v \div w)$ or if there are many such lines.

Let $(u' \div v' \div w')$ denote one of such lines. The quantities u', v', w' are subject to two conditions. First, since $(u' \div v' \div w')$ passes through the point (x_0, y_0) ,

$$u'x_0 + v'y_0 + w' = 0 \quad (*)$$

Secondly, the system of the two equations

$$\left. \begin{aligned} u'x + v'y + w' &= 0 \\ ux + vy + w &= 0 \end{aligned} \right\} \quad (**)$$

must be inconsistent since the two lines $(u' \div v' \div w')$ and $(u \div v \div w)$ have no common points. In case the system $(**)$ is inconsistent, then necessarily $u' \div u = v' \div v$, or, denoting each of the equal ratios by μ ,

$$u' = \mu u, \quad v' = \mu v$$

Then $(*)$ immediately yields

$$w' = -\mu (ux_0 + vy_0)$$

whence

$$u' \div v' \div w' = u \div v \div - (ux_0 + vy_0)$$

Accordingly, the ratios $u' \div v' \div w'$ are completely determined, that is, there is just one line passing through (x_0, y_0) and parallel to an arbitrarily given line $(u \div v \div w)$.

Thus, for our realization the parallel properties satisfy Axiom V.

In this way, we have found a concrete realization of the system of the axioms of Groups I—V; accordingly, this system is consistent. Since this realization is based on the notion of a real number, the result thus obtained is conditional and may be formulated as follows:

The system of Axioms I-V is consistent if the arithmetic of real numbers is consistent.

The proof of the fact that arithmetic is consistent falls outside the scope of the foundations of geometry and this question will not be discussed here.

In conclusion, observe that all the relationships employed in the present subsection occur in analytic geometry while using the orthogonal Cartesian system of coordinates. Therefore sometimes the above realization will be referred to as the *Cartesian realization*.

Setting the task of constructing a concrete realization of Hilbert's axioms, we chose the objects of realization from arithmetic and, verifying axioms step by step, reached the conclusion that all the definitions given by us obey these axioms. No reference to geometrical visualization was made because of the purely arithmetical nature of objects, so the above investigation turned out to be cumbersome. The investigation was carried out in full detail, since it was very important for us to be able to assert the consistency of Hilbert's axioms (more precisely, it enabled us to reduce this question to the consistency of arithmetic).

What is more, as the reader will find in the next few pages, some modifications of arithmetical realization enable us to solve a number of questions regarding the independence of the axioms of Groups I-V.

4.3. Proof of independence of some axioms of Euclidean geometry

§ 72. In § 69 the minimal problem was noted to be one of the crucial problems of the axiomatic theory. In order to solve this problem completely, each requirement of the accepted axioms must be shown to be independent of the others, that is, that the number of axioms cannot be curtailed. Such an investigation would require much time and would not be appropriate for this book. We shall confine ourselves to the proof of the independence of some of the axioms of Groups I-V from the remaining axioms of this system.

First, it may be asserted that Axiom V, the parallel axiom, is not a consequence of the axioms of Groups I-IV. The question concerning the independence of Axiom V was already settled earlier and will not be discussed further. We shall show that the axioms of Group IV, the axioms of continuity, are independent.

In the first place, it will be shown that Cantor's axiom IV.2 does not follow from the remaining axioms (including Archimedes' axiom IV.1). In the light of the general principle of such proofs (§ 69), we must construct a set of objects and define mutual relationships between these objects so that these relationships satisfy all the axioms except Cantor's axiom.

To this end, following Hilbert, we use an infinite set of numbers Ω obtained from rational numbers by repeated applications of operations of addition, subtraction, multiplication, division and, apart from these, the fifth operation $\sqrt[3]{1 + \omega^2}$ where ω is a number obtained from these operations. Evidently, the set Ω has the following properties: if ω_1 and ω_2 are any two numbers in Ω , then the numbers $\omega_1 + \omega_2$, $\omega_1 - \omega_2$, $\omega_1\omega_2$, ω_1/ω_2 ($\omega_2 \neq 0$) and $\sqrt[3]{\omega_1^2 + \omega_2^2} = \pm \omega_1 \sqrt[3]{1 + (\omega_2/\omega_1)^2}$ are also in Ω .

Now we define the geometric objects: any pair (x, y) of numbers belonging to Ω will be called a point; the ratio $(u \div v \div w)$ of three numbers of Ω , assuming that at least one of the numbers u or v is not zero, will be called a line. All mutual relationships (incidence relation between a point and a line, congruence, etc.) of objects are defined in exactly the same manner as in § 71, where the Cartesian realization of the axioms of Groups I-V was constructed. However, the coefficients in the formulas for the orthogonal transformations are now to be taken from Ω . In verifying the axioms of Groups I-III and V, we used only comparison of numbers according to their magnitude, arithmetical operations (addition, subtraction, multiplication and division) and the operation of taking the square root of the sum of squares of two numbers (the latter operation was employed in normalizing the parameters of a half line). As already noted, these operations as applied to numbers in Ω yield numbers belonging to Ω . Therefore, the conclusions drawn in the course of verifying the axioms of Groups I-III and V for the Cartesian realization remain valid when the numbers are taken from Ω . It may be thus asserted that the axioms of Groups I-III and V are fulfilled for the new realization in question.

The matter is different with the axioms of Group IV. We verify separately Archimedes' axiom IV.1 and Cantor's axiom IV.2. First, note that by means of congruence (in our realization, by means of an orthogonal transformation) every half line may be superposed on a given half line. Therefore, it suffices to verify Archimedes' axiom on any one line. The x -axis, that is, the line containing points

of the form $(x, 0)$, is the most convenient for this purpose. Evidently, the points $A_0(0, 0)$, $A_1(a, 0)$, $A_2(2a, 0)$, $\dots$, $A_n(na, 0)$, $\dots$, $a > 0$, determine a sequence of segments $A_0A_1 \equiv A_1A_2 \equiv \dots \equiv A_nA_{n+1} \equiv \dots$ congruent to each other. Indeed, there is an orthogonal transformation

$$x' = x + a$$

$$y' = y$$

mapping each of these segments into its neighbouring segment on the right side. Let $B(b, 0)$ be any point subject only to the condition $b > a$. In order for Archimedes' axiom to hold for the realization in question, there must be a positive integer n such that B lies between A_0 and A_n . The points A_0, B, A_n will lie in this order, provided $na > b$. But Archimedes' proposition is true in arithmetic: whatever the numbers $a > 0$, $b > 0$, $b > a$, there is a positive integer n such that $na > b$. Thus Archimedes' proposition also holds for the realization in question.

In contrast, Cantor's axiom fails to be valid in this realization. Indeed, if Cantor's axiom IV.2, besides the axioms of Groups I-III and Axioms IV, 1 and V, also holds for the system of points and lines, then it can be shown that in this system there is always a segment whose length is equal to a preassigned number (see Chapter 2, § 21, Theorem 35); however, for our realization the lengths of segments are expressed only by numbers of Ω .

Thus we may conclude that there is a system of objects mutual relationships among which satisfy the axioms of Groups I-III and Axioms IV.1 and V but not Cantor's axiom IV.2. In other words, Cantor's axiom is not a consequence of the remaining axioms of elementary geometry.

If one notes that the set Ω is denumerable, then the result obtained can be expressed in an alternative form: the uncountability of the set of elements of geometry cannot be established by means of the axioms of Groups I-III and Axioms IV.1 and V, only, without Cantor's axiom.

§ 73. Now we shall show that the first axiom of Group IV, that is, Archimedes' axiom, does not depend on the axioms of the remaining Groups I-III and V either.

To do this, we should find a realization for the axioms of Groups I-III and V in which Archimedes' proposition does not hold. Such a realization will be pointed out below; it is based, like the realization just considered, on arithmetic, but on arithmetic in a generalized sense which concerns a non-Archimedean system of numbers.

In order to make the following discussion completely clear, the fundamental propositions regarding properties of real numbers are listed (these propositions will be referred to as axioms of arithmetic).

(1) There is an operation "addition" which out of a number a and a number b yields a certain number c ; in symbols,

$$a + b = c$$

(2) There is another operation "multiplication" which out of a number a and a number b yields a definite number d ; in symbols,

$$ab = d$$

(3) If a, b, c are any three numbers, the following relations hold:

$$a + (b + c) = (a + b) + c$$

$$a + b = b + a$$

$$a(bc) = (ab)c$$

$$a(b + c) = ab + ac$$

$$ab = ba$$

(4—definition of difference) If a and b are two given numbers, then there is only one number x such that $a + x = b$.

Axioms 3 and 4 imply that there is only one number, called *zero* and denoted 0, such that

$$a + 0 = a$$

for every number a .

(5—definition of quotient) If a and b are two given numbers and $a \neq 0$, then there is only one number x such that $ax = b$.

Axioms 3 and 5 imply that there is only one number, called *identity* and denoted 1, such that

$$a \cdot 1 = a$$

(6—property of order) If a and b are two distinct numbers, then one of them is always greater ($>$) than the other, and then the second is less ($<$) than the first. In symbols, either

$$a > b \quad \text{and} \quad b < a$$

or

$$b > a \quad \text{and} \quad a < b$$

Moreover, if $a > b$ and $b > c$, then $a > c$; if $a > b$, then $a + c > b + c$; if $a > b$ and $c > 0$, then $ac > bc$; there is no such a for which the relation $a > a$ is true.

(7—Archimedes' proposition) If a and b are any two positive numbers ($a > 0$ and $b > 0$), then the number a may be added to itself as many times as to make the sum greater than b :

$$a + \dots + a > b$$

(8—Cantor's proposition) or any other equivalent proposition.

All these axioms are applicable to the set of real numbers and arithmetical operations on them. Whether the above Propositions 1-8 constitute a complete system of axioms for arithmetic, that is, whether any theorem of arithmetic can be proved with their aid is immaterial for the present discussion. But a careful analysis of the arguments and computations employed in the course of verifying the geometric axioms for the Cartesian realization would reveal that we need only the properties of numbers which are expressed in Propositions 1-8. In view of this the notion of number may be examined from the axiomatic point of view and thus the class of objects of arithmetical realization may be widened; such a possibility is important for our investigation.

Consider a set A , the nature of whose elements is irrelevant for us. Suppose that with every pair of elements a, b of set A (b may coincide with a) an element c of the same set is associated. We shall call this association *addition* and element c the *sum of elements a and b* ; the sum will be expressed in the usual notation: $c = a + b$. Suppose further that with every pair of elements a, b of A (b may coincide with a) an element d of the same set is associated in a different way. This association will be called *multiplication* and d will be called the *product of a and b* and we shall write $d = ab$.

Finally, assume that the elements of A are situated in a definite order, that is, for any two elements a and b one of them is considered as preceding the other; we shall call the preceding element "less than" the element which follows it.

The elements of set A will be called *generalized numbers* if the operations of addition and multiplication and the order among the elements are defined in such a way that all the conditions of Propositions 1-8 are fulfilled.

Now suppose that geometric elements and relationships between them are defined in exactly the same way as in course of constructing the Cartesian realization but, instead of usual numbers, generalized numbers are taken. Obviously we obtain a certain realization for the geometric axioms of Groups I-V, no matter what is the nature of generalized numbers employed. It is perfectly clear that the realization thus constructed differs little from the Cartesian realization. Indeed, although in constructing geometric objects elements of arbitrary nature are used, the operations on these elements obey the usual rules of arithmetic.

However, a further generalization of the notion of a number is possible. This proves to be useful and enables us to solve the following problem: we want to show that the Archimedes' axiom is independent of the axioms of Groups I-III and V. Suppose that a set A is given on whose elements the operations of addition and multiplication are defined and a certain order is established; the set A will be called a *non-Archimedean system of numbers (generalized)*

if the elements of A satisfy Propositions 1-6 but Archimedes' proposition 7 does not hold. A non-Archimedean system is described below.

Consider the set of all possible rational functions of the form

$$\omega(t) = \frac{a_0 t^n + a_1 t^{n-1} + \dots + a_n}{b_0 t^m + b_1 t^{m-1} + \dots + b_m}$$

with real coefficients a_h and b_h . To this set we adjoin all the functions obtained from rational functions by repeated applications of the operations of addition, subtraction, multiplication, division and the fifth operation $\sqrt{1 + \omega^2(t)}$, where $\omega(t)$ is the result of these operations. Denote the set of functions thus obtained by $\Omega(t)$. The set $\Omega(t)$ evidently contains all rational functions and, in particular, the functions of the form $\omega(t) = \text{const.}$, that is, functions whose values remain a fixed number as t varies.

We wish to treat the elements of $\Omega(t)$ as generalized numbers. To do so, the operations of addition and multiplication must be defined first. Take two functions $a(t)$ and $b(t)$ of $\Omega(t)$; having agreed to treat them as generalized numbers, we shall simply write a and b instead of $a(t)$ and $b(t)$. Clearly, $a(t) + b(t) = c(t)$ is a function belonging to $\Omega(t)$, and so is $a(t)b(t) = d(t)$. Thus $c(t)$ and $d(t)$ are generalized numbers c and d ; the former will be called the sum of a and b and the latter their product. Since for each t the operations $a(t) + b(t)$ and $a(t)b(t)$ obey the usual rules of arithmetic, the operations of addition and multiplication just defined meet all the requirements of Propositions 1-5. The zero in the system of generalized numbers is the function which is identically equal to the usual zero and the generalized identity is the function which is identically equal to the usual identity.

Since the given system of generalized numbers satisfies Propositions 1-5, all four arithmetical operations are defined. Observe that the operation $\sqrt{a^2 + b^2}$ is also defined in our system. Indeed, if $a(t)$ and $b(t)$ are two functions in $\Omega(t)$, then

$$\sqrt{a^2(t) + b^2(t)} = \pm a(t) \sqrt{1 + \left(\frac{b(t)}{a(t)}\right)^2}$$

is also a function in $\Omega(t)$. This function may be treated as a generalized number $\sqrt{a^2 + b^2}$ which is defined by the two given numbers a and b .

Now order in the set $\Omega(t)$ will be defined. Let $\omega(t)$ be any function in $\Omega(t)$ which represents number ω in our system. If $\omega \neq 0$, that is, if $\omega(t)$ is not identically zero, then, with sufficiently large t^* , for any $t > t^*$ the function $\omega(t)$ retains a definite sign.* If $\omega(t) > 0$

* This is the consequence of the fact that $\omega(t)$ is an algebraic function (any algebraic function changes its sign a finite number of times).

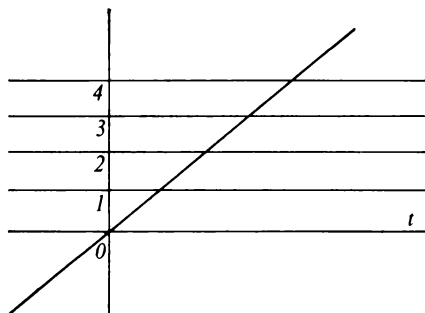


Figure 86

for $t > t^*$, the generalized number ω will be considered positive, $\omega > 0$. If, however, $\omega(t) < 0$ for $t > t^*$, it will be assumed that $\omega < 0$. Dividing all the generalized numbers (except zero) in this manner into positive and negative numbers, we compare the numbers according to their magnitude in the usual way: $a > b$ if $a - b > 0$.

It can be easily verified that all the conditions of Proposition 6 are satisfied.

At the same time, in our system of generalized numbers Proposition 7 does not hold; accordingly, the system is non-Archimedean. To show this, it is profitable to formulate the above definition of the inequality $a > b$ in terms of the following geometrical form: $a > b$ if the graph of the function $a(t)$ rises, as $t \rightarrow +\infty$, above the graph of the function $b(t)$. As already noted, among the elements of set $\Omega(t)$ there are functions which retain, as t varies, a constant value: $\omega(t) \equiv c$. The graphs of such functions are lines parallel to the t -axis. From our point of view, each function $\omega(t) \equiv c$ is a generalized number; we denote it simply by c so that if number 10 or 20 is written this means that the function $\omega(t)$ is identically equal to 10 or 20. Consider the two functions $a(t) \equiv 1$ and $b(t) \equiv t$ which are contained in $\Omega(t)$ and may therefore be considered as generalized numbers a and b . If the number a is added to itself n times, the sum belongs to the set $\Omega(t)$ and the graph of this sum is a line parallel to the t -axis and is situated in the positive half plane at a distance n from the t -axis. The graph of $b(t) \equiv t$ is a bisector of the first coordinate angle. But, as $t \rightarrow +\infty$, the graph of function $b(t) \equiv t$ rises higher than any line parallel to the t -axis (Fig. 86). This implies that no matter how many times number a is added to itself, always

$$a + \dots + a < b.$$

Thus in our system of generalized numbers Archimedes' proposition does not hold.

Now it is a simple matter to construct a system of geometrical objects for which the axioms of Groups I-III and V hold but Archimedes' axiom does not.

By a point we shall mean a pair (x, y) of numbers of the non-Archimedean system $\Omega(t)$, by a line the ratio $(u \div v \div w)$ of three numbers u, v, w of $\Omega(t)$ subject to the only condition that one of the two numbers u or v is not zero. All the mutual relationships between geometric objects are defined in exactly the same manner as in the Cartesian realization for Hilbert's axioms. In the system $\Omega(t)$, the operations of addition and multiplication of elements as well as the relationships "greater than" and "less than" are defined in accordance with axioms of arithmetic 1-6. Moreover, the operation $\sqrt{a^2 + b^2}$ is defined for any two elements a and b . Therefore, all the arguments and computations used in the verification of the axioms of Groups I-III and V for the Cartesian realization may be fully repeated when generalized numbers from $\Omega(t)$ are being used instead of ordinary numbers. Therefore, for the realization thus constructed, all the axioms of Groups I-III and V are valid; at the same time, since the set $\Omega(t)$ is non-Archimedean, Archimedes' axiom IV.1 does not hold in this system. This implies that Archimedes' axiom is independent of the axioms of Groups I-III and V.

Summing up the above discussion, the following propositions may be stated:

Archimedes' Axiom IV.1 cannot be proved with the aid of the axioms of Groups I-III and V.

Cantor's axiom IV.2 cannot be proved with the aid of the axioms of Groups I-III and Axioms IV.1 and V.

It is natural to raise the question as to whether Archimedes' axiom can be proved by means of the remaining axioms, including Cantor's axiom. The answer is negative. To demonstrate this, one must construct a non-Archimedean system of numbers for which Cantor's axiom holds. An example of such a system is given below*.

Let us call a number any power series

$$a_0 t^n + a_1 t^{n+1} + a_2 t^{n+2} + \dots$$

where $a_0, a_1, a_2, \dots$ are any ordinary real numbers, and n any ordinary integer (positive, negative or zero). The ordinary real numbers are included in this system as series of the type

$$a + 0 \cdot t + 0 \cdot t^2 + \dots$$

By zero we mean the series

$$0 + 0 \cdot t + 0 \cdot t^2 + \dots$$

* This example was pointed out to me by A.N. Kolmogorov.

If the number

$$a_0 t^n + a_1 t^{n+1} + \dots$$

is not zero, then the ordinary number a_0 is also assumed not to be zero.

Let the operations of addition and multiplication coincide with the formal operations of addition and multiplication of power series (that is, by the sum of two numbers of our system represented by power series we shall mean the power series obtained by adding like members of the series; by the product of two numbers of our system represented by certain power series we mean the power series obtained by multiplying the two series termwise and then arranging the terms in ascending powers of t).

It may be easily seen that Axioms 1-5 of arithmetic are fulfilled. What is more, defined in our system is the operation $\sqrt{1 + \omega^2}$, where ω is any number of the system. (Seeking the quotient $x = b/a$, $a \neq 0$, leads to successive determination of unknown coefficients of the series for x by comparing the two sides of the equation

$$ax = b$$

The number

$$x = \sqrt{1 + \omega^2}$$

is similarly found by means of the equation

$$x^2 = 1 + \omega^2$$

Now we shall introduce order in our system of numbers. We shall call the number

$$a_0 t^n + a_1 t^{n+1} + \dots$$

$a_0 \neq 0$, positive (greater than zero) if $a_0 > 0$, negative (less than zero) if $a_0 < 0$. If a and b are two numbers of our system, then $a > b$ if $a - b > 0$ and $a < b$ if $a - b < 0$. This order satisfies Axiom 6 of arithmetic.

We will now show that Cantor's axiom holds in our system. Let a monotonically increasing sequence of numbers of our system

$$a^{(m)} = a_0^{(m)} t^{p_m} + \dots, \quad m = 1, 2, \dots$$

and a monotonically decreasing sequence

$$b^{(m)} = b_0^{(m)} t^{q_m} + \dots, \quad m = 1, 2, \dots$$

be given, where (i) any member of sequence $a^{(m)}$ is less than every member of sequence $b^{(m)}$, and (ii) for any positive number ε (of our system) there is a number m such that

$$b^{(m)} - a^{(m)} < \varepsilon$$

We show that there is a (unique) number in our system which lies inside each of the segments $(a^{(m)}, b^{(m)})$.

First note that sequences of ordinary numbers p_m and q_m ($m = 1, 2, \dots$) are bounded from below. In fact, if the ordinary numbers p_m contain numbers situated as far as possible on the left of zero, then from the powers of p_m a subsequence may be chosen which monotonically converges to $-\infty$; the corresponding initial coefficients must be positive; otherwise the condition that the numbers of sequence $a^{(m)}$ increase monotonically would be violated. In this case one of the members of $a^{(m)}$ would be greater than some member of $b^{(m)}$, which is impossible.

In the same way it can be proved that the numbers q_m are bounded from below. We may thus assume that all the series representing $a^{(m)}$ and $b^{(m)}$ start with the like power (allowing for this proof the value zero for the initial coefficients).

We rewrite these series in the form

$$\begin{aligned} a^{(m)} &= a_0^{(m)}t^n + a_1^{(m)}t^{n+1} + \dots \\ b^{(m)} &= b_0^{(m)}t^n + b_1^{(m)}t^{n+1} + \dots \end{aligned}$$

It is easily seen that starting with a certain number $m = m_1$, the nonnegative difference $b_0^{(m)} - a_0^{(m)}$ becomes zero. Indeed, since the sequence $a^{(m)}$ is increasing, the sequence of (ordinary) numbers $a_0^{(m)}$ must be nondecreasing. Similarly, since the sequence $b^{(m)}$ decreases, the sequence of (ordinary) numbers $b_0^{(m)}$ must be non-increasing. Thus the difference of (ordinary) numbers $b_0^{(m)} - a_0^{(m)}$ cannot increase. This means that $b_0^{(m)} - a_0^{(m)}$ is either always positive or is zero for a certain number m and then remains zero for the successive numbers. Suppose that always

$$b_0^{(m)} - a_0^{(m)} > 0$$

Taking in our system the positive number

$$\varepsilon = t^{n+1} + \dots$$

we have

$$b^{(m)} - a^{(m)} > \varepsilon$$

for an arbitrary number m , which contradicts the hypothesis. Thus the difference $b^{(m)} - a^{(m)}$ cannot remain positive.

Accordingly, starting with a certain $m = m_1$, we have

$$b_0^{(m)} - a_0^{(m)} = 0$$

Since the sequence of ordinary numbers $a_0^{(m)}$ is monotonically nondecreasing and the sequence $b_0^{(m)}$ is monotonically nonincreasing, it follows that starting with a certain $m = m_1$ the numbers $a_0^{(m)}$ and $b_0^{(m)}$ remain unchanged and coincide; set $a_0^{(m_1)} = b_0^{(m_1)} = d_0$.

Then

$$a_0^{(m)} \leq d_0 \leq b_0^{(m)}$$

If $m \geq m_1$, analogous arguments are applicable to the sequences of ordinary numbers $a_1^{(m)}$, $b_1^{(m)}$; in exactly the same manner it is established that there is a number d_1 such that

$$a_1^{(m)} \leq d_1 \leq b_1^{(m)}$$

for $m \geq m_1$; moreover, starting with a certain $m = m_2$ ($m_2 \geq m_1$) the difference $b_1^{(m)} - a_1^{(m)}$ vanishes, etc.

The number $d = d_0 t^n + d_1 t^{n+1} + \dots$ lies inside each of the segments $(a^{(m)}, b^{(m)})$. This shows that Cantor's axiom holds for our system (that number d is unique follows immediately from the second condition in Cantor's axiom).

Archimedes' axiom does not hold for the system under consideration. Indeed, taking two positive numbers

$$\begin{aligned} a &= t + 0 \cdot t^2 + \dots \\ b &= t^2 + 0 \cdot t^3 + \dots \end{aligned}$$

we have

$$nb < a$$

for any natural number n , that is, Archimedes' axiom is violated.

In the arithmetical realization for Hilbert's axioms based on the system of numbers just described, Cantor's proposition holds (as do all the axioms of Groups I-III and V), but Archimedes' proposition does not hold. Accordingly, one may assert that

Archimedes' axiom IV.1 cannot be proved by means of the axioms of Groups I-III and Axioms IV.2 and V.

Thus both axioms constituting the fourth group are relevant.

A geometric system developed on the basis of the axioms of Groups I-III (or the axioms of Groups I-III and V) in which Archimedes' principle does not hold is known as *non-Archimedean*. In non-Archimedean geometry the process of measurement is not applicable to an arbitrary segment; what is more, many propositions of this geometry differ from propositions of both Euclidean and Lobachevskian geometries. This is not surprising, since Archimedes' axiom is needed in proving many theorems. In particular, not valid in non-Archimedean geometry are the results of Legendre which establish a connection between the parallel axiom and the proposition concerning the angle-sum of a triangle (for details see D. Hilbert: *Foundations of Geometry*, 2nd ed., Open Court, La Salle, Ill., 1971).

4.4. Axiom of completeness

§ 74. In Chapter 2, the properties of continuity are expressed by two axioms: Archimedes' axiom IV.1 and Cantor's axiom IV.2. In his *Foundations of Geometry* Hilbert lists the first axiom of continuity which is the same as the one mentioned in Chapter 2, but the second axiom differs from that of Cantor's, and Hilbert calls it the *axiom of completeness*. This axiom may be formulated in the following way.

The elements (points, lines, planes) of geometry constitute a system of objects which is such that, if the conditions of all previously accepted axioms are satisfied, this system cannot be extended, that is, the system of points, lines, and planes is such that no new points, lines and planes can be added such that all the axioms of Groups I-III and Axioms IV.1 and V continue to hold in the extended new system.

The continuance of the validity of all the axioms in the above proposition should be understood in the sense that for the extended system all the axioms are satisfied as before so that, in particular, the existing relationships between the elements, their order, congruence of segments and angles, etc. are not disturbed; for example, if a point lay between two points before extension, it should continue to do so after the extension; congruent segments and angles continue as such after the extension of the system. In order to make clear the axiom of completeness of systems of geometric elements, let us compare the two realizations for axioms examined in §§ 71 and 72.

The first of these is the Cartesian realization for which all the axioms, without any exception, are satisfied. In the Cartesian realization a *point* means a pair of any real numbers (x, y) , a *line* means the ratio $(u \div v \div w)$ of three numbers which are chosen subject to the only condition that either u or v is not zero. The mutual relationships between the objects are expressed in terms of arithmetical formulas which will not be repeated here.

The realization constructed in § 72 is totally analogous to the Cartesian realization. Here also a point is the pair of real numbers, a line is the ratio of three numbers; the mutual relationships between the objects are expressed by exactly the same formulas as in the Cartesian realization. But in contrast to the Cartesian realization, the objects of this realization are not constructed from all real numbers but only from those numbers which belong to the set Ω which was described earlier in detail. Thus the aggregate of the objects of realization in § 72 is a part of the aggregate of the objects of the Cartesian realization, but in both realizations all the axioms of Groups I-III and Axioms IV.1 and V are fulfilled.

Consider a set of objects defined by means of numbers belonging to Ω as the initially given set and the set of objects of the Cartesian realization as having been obtained by completing the first set. Since the mutual relationships between the objects of the two realizations in question are expressed by the same arithmetical formulas (only in the first case these formulas concern all real numbers and in the second they concern only real numbers from a certain set), all the mutual relationships between the initial objects remain valid for the objects of the completed set. For example, if A, B, C, D are four points of the initial set and $AB \equiv CD$, then $AB \equiv CD$ continues to hold after the addition of new elements. What is more, the relationships between the new objects and the initial objects as well as the relationships between the new objects are completely defined, and all the initial axioms are satisfied.

Therefore the set of objects defined in the manner described with the aid of numbers in Ω possesses the very extension denied by the axiom of completeness. In other words, this aggregate of objects does not satisfy the completeness condition.

There is an infinite number of such systems. To obtain these systems, it is enough to modify the construction of the set from which the required numbers are drawn. For example, in the construction of objects one may use, instead of Ω , the set of numbers expressed in terms of radicals, a wider set of algebraic numbers, etc. Among the arithmetical realizations thus obtained only the Cartesian realization (based on the set of all real numbers) satisfies the completeness axiom. In order to show this, it should be noted first that among all arithmetical realizations only the Cartesian realization satisfies Cantor's axiom (or Dedekind's condition) and, secondly, that Cantor's axiom together with the remaining axioms implies the completeness axiom. The first fact needs no proof because, as shown earlier, Cantor's axiom holds for the Cartesian realization and, on the other hand, among all the arithmetical realizations Cantor's axiom is satisfied only for the Cartesian realization since Cantor's axiom (or Dedekind's principle) is valid only for a set containing at least one number.

The second fact will be proved. It will be shown that not only does Cantor's axiom combined with the remaining axioms imply the completeness axiom but also that, conversely, Cantor's axiom follows by adjoining the completeness axiom to the remaining axioms. The stated assertion may be formulated in detail in the following manner:

If a system of elements of geometry satisfies the axioms of Groups I-V, then this system cannot be extended to a system in which the completeness axiom holds, that is, the completeness axiom is a consequence of the axioms of Groups I-V. If the system of elements of geometry satisfies

the axioms of Groups I-III, Axioms IV.1 and V and the completeness condition, then Cantor's axiom holds in this system, that is, Cantor's axiom follows from the axioms of Groups I-III, Axioms IV.1 and V and the completeness axiom.

We start with the proof of the first part of this proposition. Let Σ denote the set of elements of geometry, that is, a system of points, lines and planes whose mutual relationships satisfy the axioms of Groups I-V, and let Σ' denote the completion of Σ by introducing new elements satisfying the conditions mentioned in the completeness axiom. The mutual relationships between the elements of the extended system also satisfy the axioms of Groups I-V. In § 22 it was shown that the elements of geometry may be arithmetized by means of the axioms of Groups I-III and Axiom IV.1 such that every point has as its coordinates a completely defined triple of numbers (x, y, z) and that the same triple cannot be assigned to distinct points.

Let a coordinate system be introduced for elements of Σ' , where as the unit of measurement a segment with endpoints lying in Σ is chosen. Suppose that Σ' contains points that are not in Σ . Let M' be one of such points and let (x, y, z) be its coordinates. By hypothesis, the original elements of Σ satisfy the axioms of Groups I-V; therefore, taking note of Theorem 35 of § 21, we see that among the points of Σ there is always one which has the preassigned coordinates. Let M be a point of Σ with coordinates (x, y, z) . Since M' is not contained in Σ , it follows that M' and M cannot coincide. Thus the triple (x, y, z) is assigned to two distinct points M and M' . This contradiction demonstrates that Σ' does not contain more points than are already contained in Σ .

Assume that Σ' has lines not contained in Σ . Take one such line a' . According to Axiom I.3, line a' contains at least two points A and B . These points are in Σ because Σ' does not contain new points. But Σ realizes Axiom I.5 in its own right; therefore the pair A, B determines a certain line a contained in Σ . Lines a' and a cannot coincide, since a' is not contained in Σ . Thus A and B determine two distinct lines, a fact which violates Axiom I.2. The contradiction shows that Σ' does not contain new lines apart from those already in Σ . Similarly, it may be shown that Σ' does not contain new planes. Hence it may be concluded that Σ cannot be extended, that is, it satisfies the completeness axiom.

Now we prove the second part of the proposition, and for convenience confine ourselves to the case of plane geometry. Assume that Σ now denotes the set of points and lines satisfying the axioms of Groups I-III and Axioms IV.1 and V. The validity of Cantor's axiom is not assumed beforehand; instead it is assumed that Σ satisfies the completeness axiom.

We must derive Cantor's axiom from the accepted axioms. To do this, we introduce in Σ a coordinate system by choosing, as was done in § 22, two mutually perpendicular lines and a segment as the measuring scale. Then to every point there is assigned a pair of coordinates (x, y) . If we could apply Cantor's axiom, then, by Theorem 35 of § 21, it could be concluded that the coordinates of points of Σ exhaust all possible pairs of numbers. But now we try to establish this result by using the completeness axiom in place of Cantor's axiom. Having done this, the truth of Cantor's principle for Σ may be directly established.

For points and lines of the set Σ all the theorems of Euclidean geometry are valid except possibly those which concern properties of continuity (because Axiom IV.2 is not accepted now). In any case, the chosen coordinate system will have all the basic properties of the Cartesian coordinate system. In this system a line is determined by a first degree equation

$$ux + vy + w = 0$$

and thus to every line there corresponds the ratio of three numbers $(u \div v \div w)$. Employing the usual tool of analytical geometry, all the mutual relationships between points and lines of Σ required by the axioms of Groups I-III and Axioms IV.1 and V may be characterized by arithmetical formulas containing coordinates x, y of points and coefficients u, v, w in equations of lines. Evidently, the forms of these formulas are identical to those obtained in describing the Cartesian realization of geometric axioms.

Suppose now that there are pairs of numbers (x, y) which are not assigned as coordinates to any point of Σ and ratios $(u \div v \div w)$ which are not assigned as ratios of coefficients of any equations representing lines belonging to Σ . As will be shown now, in this case the set of elements of geometry Σ may be extended such that the completeness axiom is satisfied.

Now, to the set Σ let us add new points and lines defining them in the following manner: a new point is any pair of numbers (x, y) which does not happen to be a pair of coordinates for any point of Σ , and a new line is the ratio of any three numbers $(u \div v \div w)$, where either u or v is not zero, such that u, v, w are not coefficients of any equation representing a line of Σ . The set of points and lines thus extended will be denoted by Σ' . The points and lines of Σ' are uniquely determined by the pairs of numbers (x, y) and by the ratios of three numbers $(u \div v \div w)$, respectively, and as arithmetical representatives of the elements of Σ all possible combinations of all real numbers will be found.

All the mutual relationships between the elements of Σ' will be defined by exactly the same arithmetical formulas as in the description of the Cartesian realization. Then, evidently, the points

and lines of Σ' satisfy all the axioms of Groups I-III and Axioms IV.1 and V because these are satisfied in the Cartesian realization. Moreover, it is clear from the previous remarks that those elements of Σ' which also belong to Σ retain the same mutual relationship as before the extension. In fact, before and after extensions these relationships are expressed by the same arithmetical formulas. Thus the set Σ has been extended in a way such that the conditions mentioned in the completeness axiom are satisfied. But since we have accepted the truth of the completeness axiom, and it rules out any such extension, we must conclude that all the possible pairs of real numbers have been exhausted as coordinates (x, y) of points of the given set Σ . In that case every line of Σ may be treated as the number axis whose points represent real numbers. Since Cantor's principle holds for the set of real numbers, it must also hold for the set of points of every line in Σ .

Summing up the discussion, we may state the following:

We have shown that assuming the validity of the axioms of Group I-III and Axioms IV.1 and V, the continuity of a set of elements may be demonstrated in two equivalent ways: accepting as an axiom either Cantor's proposition of the nested system of segments or Hilbert's proposition on the completeness of the system of geometric elements. If one of these propositions is accepted without proof, the other may be proved as a theorem.

We note one more interesting feature of the completeness axiom. This axiom cannot be retained in the list of axioms if Archimedes' axiom is omitted. This is due to the fact that without satisfying Archimedes' axiom, the system of geometric elements may be completed by new elements such that the mutual relationships between the initial elements continue to be valid. Therefore without Archimedes' axiom the completeness axiom yields a contradiction. Thus in Hilbert's exposition both the axioms of continuity are connected organically, the first of these prepares being the ground for the continuity requirement, while the second expresses this requirement by the completeness condition.

4.5. The completeness of axioms of Euclidean geometry

§ 75. In § 69, three fundamental problems of axiomatic theory were pointed out: the problem of consistency, the minimal problem and the problem of completeness. These problems arise in a natural way in the investigation of any axiomatic system. The first two problems for Hilbert's system of axioms were disposed of in the previous articles. Now we shall deal with the third problem.

First we will try to explain its meaning. Let us review the situation which prevailed, due to the development in geometry, in the second half of the nineteenth century. By this period all the main geometrical

disciplines had completely taken their shapes and the main question was their axiomatic substantiation. It was already clear that Euclid's ancient system of axioms could not serve as a base for the logical development of geometry. What was required was a complete system of axioms (and definitions), that is, a system containing all propositions which could be used in proving theorems of elementary geometry without referring to any diagrams. Our discussion in Chapter 2 convinced us that, by using Hilbert's axioms, theorems can be proved in a strictly logical manner.

A natural question arises: How is the completeness of Hilbert's system of axioms to be precisely understood? Of course, one may think that the completeness of Hilbert's system was discovered by analyzing all the theorems of geometry known prior to, say, 1900. Such an answer would be satisfactory only if elementary geometry was to be regarded as a discipline whose growth is completed. But, although from the historical point of view the problem of substantiating elementary geometry was being solved at the time when elementary geometry was sufficiently developed, from a purely mathematical point of view this problem cannot be posed by treating geometry as enclosed in some sort of frame because the number of possible theorems of geometry has no bound. Therefore, we try to define the notion of completeness in such a way that it could be applied to a given system of axioms irrespective of the degree of development of that geometry which is to be substantiated by these axioms.

Suppose that the axioms of a given system are realized on two distinct sets of objects. Two realizations of the axioms will be called *isomorphic* if a one-to-one correspondence may be established between the objects of these sets such that the corresponding objects have identical mutual relationships. (For example, if a point A and line a of the first realization correspond to point A' and line a' of the second and if point A lies on a , then point A' lies on a' ; if the segments AB and CD of the first realization correspond to segments $A'B'$ and $C'D'$ of the second and if $AB \equiv CD$, then $A'B' \equiv C'D'$, and so on. The relationships "lies on", "between", "congruent to" must be understood for each realization in the specific sense of the realization in question.)

We explain this definition by examples.

It was shown in § 46 that the axioms of Lobachevskian plane geometry are realized on any equidistant surface. Let us consider a system of equidistant surfaces with a common base σ , and let Σ and Σ' be any two equidistant surfaces of this system. A point of Σ will be said to correspond to a point of Σ' if they lie on one ray orthogonal to σ ; two equidistant curves of Σ and Σ' will be said to correspond if they lie in one plane orthogonal to the base σ . In this way a correspondence between the objects of the realizations of

Lobachevskian geometry on Σ and Σ' is established. This correspondence is evidently an isomorphism.

Let us treat the first three axioms in Hilbert's Group I as an independent system of axioms. A realization for this system will be obtained if by points we mean three vertices of a triangle while by lines the sides of the triangle. The conditions of Axioms I.1-I.3 are satisfied although there are only six objects in all (this realization reminds us of the realization described in § 70, but the present one is simpler because now only a part of Group I axioms are taken into account). On the other hand, let us recall the arithmetical realizations of Hilbert's axioms presented in § 71 and § 72; here they will be treated as a realization only of Axioms I.1-I.3 (that is, the fact that for these realizations other axioms are also valid is of no interest for the present discussion). All these realizations are nonisomorphic. Indeed, in the first case Axioms I.1-I.3 are realized only on six objects whereas in § 72 they are realized on an infinite but countable set of objects and in § 71 the set, being infinite, is uncountable. Thus, no matter which two of the three realizations are taken, no correspondence, let alone isomorphic, between their elements can be established.

Evidently, the smaller the number of conditions imposed in the axioms of a given system the larger the number of realizations for this system. For example, a system containing only Hilbert's axioms I.1-I.3 may be realized in any one of the three mentioned ways. But if the axioms of Group II are adjoined to Axioms I.1-I.3, then the first way is no longer applicable because the axioms of Groups I-II imply that the set of geometric objects is infinite. Furthermore, the axioms of Groups I-III and Axiom IV.1 may be realized as described in § 71 or § 72. But if Axiom IV.2 is also adjoined to these axioms, then the realization described in § 72 becomes inapplicable, since Axiom IV.2 is not valid for this realization. Thus, if a system of axioms is extended by adjoining new axioms which neither depend on the old axioms nor contradict them, then the class of admissible realizations of the system is narrowed.

Now the notion of the completeness of a system of axioms may be precisely formulated.

A given system of axioms is called complete if all the realizations of this system are isomorphic.

That the system of the axioms of Groups I-V is complete is proved below.*

Consider a realization Σ for the axioms of Groups I-V. According to § 22, a coordinate system may be introduced in the set of objects called points in the realization such that to every point there corresponds a unique pair of coordinates (x, y) and to every pair of

* We again confine ourselves to the case of plane geometry.

numbers (x, y) there corresponds a unique point with coordinates (x, y) . What is more, since Axiom V (the parallel axiom) is true, the coordinate system constructed in § 22 is Cartesian. Therefore the coordinates of points located on some line are characterized by the equation

$$ux + vy + w = 0$$

Hence to the points of realization of Σ there uniquely correspond pairs of real numbers (x, y) and to lines the ratios of the form $(u \div v \div w)$.

Thus a one-to-one correspondence is obtained between the objects of realization Σ and the objects of arithmetical realization discussed in § 71. This correspondence is isomorphic. To see this, it suffices to observe that the axioms of Groups I-V of elementary geometry enable us to derive the basic Cartesian formulas expressing the mutual relationships of objects of Σ arithmetically in exactly the same manner as the relationships between the objects of the realization obtained in § 71.

It is seen that every realization of the axioms of Groups I-V is isomorphic to the Cartesian realization. But two realizations which are isomorphic to a third are evidently isomorphic to each other. Consequently, all the realizations of the axioms of Groups I-V are isomorphic to each other. Therefore it can be concluded that the system of the axioms of Groups I-V is complete.

Similar arguments can be used to establish that the system of axioms of Lobachevskian geometry is complete (the basic formulas of this geometry are to be established beforehand starting from the axioms; see §§ 216-224).

4.6. The axiomatic method in mathematics

§ 76. Up to now we examined only two concrete systems of axioms: the system of axioms of Euclidean geometry and the system of axioms of Lobachevskian geometry.

Of course, in this chapter a few systems were examined which were obtained by omitting one or several axioms from Hilbert's list, but in essence such systems, being parts of Hilbert's system, do not offer anything new.

At the same time, the general point of view on geometrical objects and geometric axioms which was achieved in the investigation of the basic problems of axiomatic theory in elementary geometry enabled us to discover the possibility of applying the axiomatic method in wider fields.

At present many branches of mathematics are based on suitably constituted systems of axioms. Such, for example, are the theory of groups, set-theoretic topology, and various branches of functional

analysis. These axioms, on which the branches just mentioned are based, take note of only some properties of the mathematical objects under study. As a rule, these properties are common to numerous classes of objects whose other properties are different. The theorems obtained from the accepted axioms are applicable to all such classes of specific mathematical objects. That the mathematical conclusions are common is one of the very important factors connected with the application of the axiomatic method.

It is important to note that at the foundation of the majority of mathematical theories lie incomplete systems of axioms. For example, the axioms of the group theory constitute an incomplete system, since there are nonisomorphic groups. Also based on an incomplete system of axioms are the spaces studied in set-theoretic topology. The fact that topology and group theory have at their foundation incomplete systems of axioms is the chief reason for their extensive applications.

If new axioms are adjoined to the axioms of topology, then the class of spaces whose elements satisfy the extended system of axioms will be narrower than the original class. Thus, by adjoining newer and newer axioms to the axioms of a topological space one may obtain one of the complete systems of axioms defining Euclidean space or Lobachevskian space or some other space. Note that the larger the number of axioms contained in a chosen system the richer the contents of the theory based on this system but the narrower the field of its applications, that is, the less general its theorems.

Part 2

Projective Geometry

Chapter 5

The Foundations of Projective Geometry

5.1. The subject matter of projective geometry

§ 77. In the first decades of the nineteenth century, simultaneously with the successful progress of the investigation on the foundations of geometry, a new branch of geometric knowledge, projective geometry, grew. The sources of its growth were graphic art and architecture. At first the field of applications of projective geometry was very narrow. But as this branch grew, it entered into various geometrical fields and at the end of the nineteenth century the investigations on the foundations of projective geometry and on the foundations of elementary geometry were united. The striking result of the union was the development within the framework of projective geometry of the deep theory which knitted in a common scheme the geometries of Euclid, Lobachevsky and Riemann.

§ 78. The famous French geometer J. V. Poncelet (1788-1867) chose for his investigation some special properties of geometrical figures which he called *projective*. What these properties are will now be explained.

Let A be any figure lying in a plane α , let β be any other plane and let O be a point of space contained neither in α nor in β (Fig. 87). The point O and a point M of A determine a line OM . The line OM intersects β at some point M' ; this point M' will be called the *projection of point M* (from the centre O on the plane β). The projections of all points of figure A on plane β constitute a figure A' , which is called the *projection of figure A* . The operation resulting in figure A' is known as the *central projection* from point O . Varying the choice of O and β , we can obtain an infinite number of figures by projecting A centrally; some of these figures are partly like A but in many respects they differ considerably. For example, by projecting a circle one may obtain an ellipse or a parabola or even a hyperbola, by projecting a right triangle one may obtain a triangle of any form, and so on. Thus many properties of figures are not carried on to its projection. A right triangle may not retain its properties under a projection, which, generally speaking, does not result in a right triangle again; the basic property of a circle expressed in its definition may also be violated as a result of projection, since by projecting a circle one may obtain, for example, an ellipse, and so on. Similarly, many quantities connected with a figure may undergo a change as a result

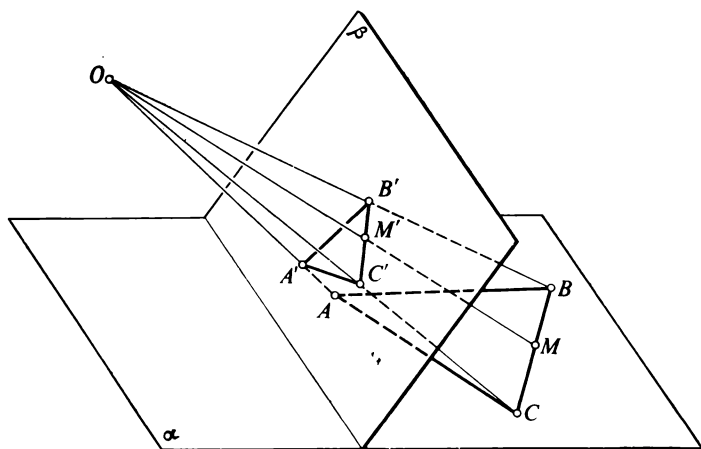


Figure 87

of projection. For example, by projecting a segment of a given length a one may obtain a segment whose length is arbitrarily large or arbitrarily small; by projecting a triangle of a given area Δ a triangle may be obtained whose area is greater or less than Δ .

On the other hand, there are properties of figures that are preserved under any projection, and with figures one may associate quantities which are preserved under any projection. Such properties and quantities are called *invariants of projection*.

The properties which remain invariant under any projection were called by Poncelet *projective properties*; he chose them as objects of investigation in projective geometry. What is more, the objects of projective geometry are the quantities that remain invariant under projection.

Examples. If the points $P_1, P_2, \dots, P_n$ of a figure A lie on a line, then their projections $P'_1, P'_2, \dots, P'_n$ also lie on a line. Accordingly, the property of rectilinear location of the points of a figure is projective. In other words, a line is an object of projective geometry.

If the points $Q_1, Q_2, \dots, Q_n$ of a figure A lie on some conic section k , then their projections $Q'_1, Q'_2, \dots, Q'_n$ also lie on a conic section k' . In other words, a conic section is an object of projective geometry. It should be noted that the properties characteristic of only a circle or an ellipse or a parabola or a hyperbola are not projective properties; therefore, in projective geometry, unlike in elementary geometry, no distinction is made between conic sections. Thus although conic sections are objects of projective geometry, their individual forms—the circles, ellipses, parabolas, and hyperbolas—

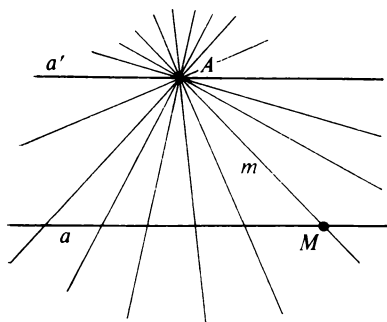


Figure 88

are not distinguished in projective geometry and therefore are not treated separately.

§ 79. The problem of investigating projective properties of figures attracted the attention of many geometers of which, besides Poncelet, mention must be made of the French mathematician M. Chasles (1793-1880) and the Swiss mathematician J. Steiner (1769-1863). They worked out a number of general questions of projective geometry in which Steiner, Chasles and their fellow geometers saw the revival of the synthetic method in geometry. In developing the synthetic method, as opposed to the analytic method, these geometers were successful in perfecting the tool of projective geometry and in applying it to various geometrical problems.

However, the principal value of projective geometry in the development of geometrical ideas is by no means determined by the number of individual cases where the methods of projective geometry prove to be more convenient than the methods of analytic geometry. As will be demonstrated, the value of projective geometry lies in its generality, which enables us to unite various geometric systems into one; in particular, it enables us to include elementary geometry in the projective scheme. But at the time of Steiner and Chasles projective geometry appeared to be a part of elementary geometry. It developed into an independent discipline towards the end of the nineteenth century.

An important factor in this development was the use in projective geometry of geometric elements at infinity. The next few subsections are specially devoted to this question.

§ 80. Let A be any point in space and a a line not through A (Fig. 88). Pass a plane α through A and a and consider all possible lines through A lying in α . These lines form a pencil in plane with centre A ; this pencil will be referred to as *pencil A* .

A correspondence may be set up between the rays of pencil A and the points of line a by associating with each point M of a the ray m

of pencil A that intersects a at M (Fig. 88). The ray m is said to project the point M .

Obviously, no matter how the point M is situated on the line a , to this point there always corresponds a definite ray. But it cannot be asserted that to every ray of pencil A there corresponds a point of line a . Namely, the ray a' of pencil A which is parallel to a does not intersect this line and therefore there is no point corresponding to the given ray. Thus the correspondence between the rays of pencil A and the points of line a is not one-to-one. This fact causes a lot of inconvenience in the study of projections. In order to salvage this situation, parallel lines are assumed to intersect at infinity. Then the ray a' , like any other ray, of the pencil A that is parallel to a will have a corresponding point which is not an ordinary point but a new object, called the *point at infinity*, of the line a .

The point at infinity of a line is also assumed to lie in every plane that passes through this line. Next, it is assumed that all parallel lines have one common point at infinity, and, accordingly, the system of parallel lines lying in one plane is called a *pencil with centre at the point at infinity*.

Note that under projection a pencil having its centre at the point at infinity may become an ordinary pencil. For example, in Fig. 89 the pencil in α having its centre at the point at infinity, S_∞ , is projected from centre O on the plane β into an ordinary pencil with centre S .

The points at infinity of nonparallel lines are treated as distinct points. Thus any plane contains infinitely many distinct points at infinity. The aggregate of all points at infinity of a plane is known

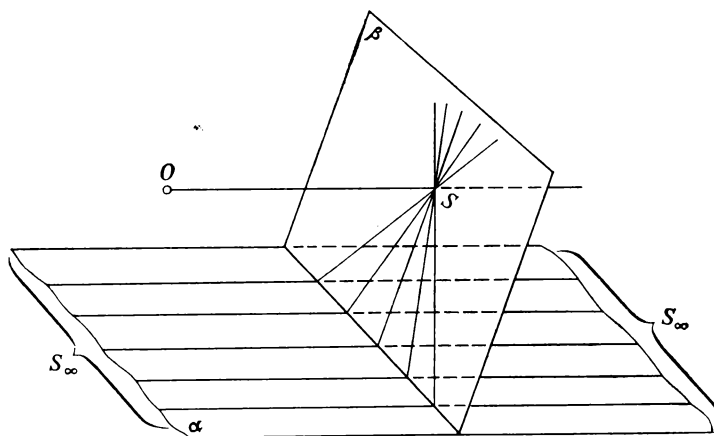


Figure 89

as the *line at infinity* of this plane. The aggregate of all points at infinity of space is called the *plane at infinity*. Such terminology may be justified by the following two factors.

(1) Two parallel planes have common points at infinity, in view of which the aggregate of points at infinity of the plane may be treated as a figure obtained by the intersection of two planes; it is then natural to treat the aggregate of points at infinity of a plane as a line at infinity.

(2) When the set of all points at infinity of the space intersects an ordinary plane, it determines a line at infinity. Therefore this set is naturally called a plane at infinity.

§ 81. All that has been said may be summed up as follows.

The set of objects of the Euclidean space is adjoined by new elements which are called "the point at infinity", "the line at infinity" and "the plane at infinity". The new elements are adjoined subject to the following conditions.

(1) A point at infinity is added to the set of points of every line; a line at infinity is added to the set of lines of every plane; a plane at infinity is added to the set of planes in space.

(2) The mutual incidence relationships among the geometric elements of the extended set must satisfy all the conditions of axioms of incidence (that is, Group I of Hilbert's axioms).

(3) The properties of mutual incidence among the geometric elements of the extended set must be such that any two planes have a common line; every line and plane have a common point; any two lines lying in one plane have a common point.

A line to which a point at infinity has been adjoined is called the *projective line*. A projective line should be viewed as a closed curve. A plane to which a line at infinity has been adjoined is called the *projective plane* and the space to which a plane at infinity has been adjoined is called the *projective space*.

§ 82. The elements at infinity are quite often introduced in elementary geometry also. But there they are essentially used only for expressing certain geometrical facts (instead of saying that lines are parallel, they are said to meet at infinity; a cylinder is treated as a cone whose vertex is at infinity, etc.). In contrast, the elements at infinity play the same role in projective geometry as the ordinary geometric elements and are an organic part of the projective space.

The reason for this difference will become clear if one compares elementary geometry with projective geometry. To a large extent, elementary geometry studies the so-called metric properties of figures, that is, properties that are connected with the measurement of geometric quantities (lengths, angles, and areas). A segment AB with ordinary endpoints may always be measured and the result is a definite number expressing the length of the segment AB . But when one of the endpoints is at infinity, the measuring process loses

its meaning because for such a segment the linear unit is applied an infinite number of times. Similarly, an angle cannot be measured if one of its sides is at infinity, and the "naive" techniques of measuring areas are inapplicable to figures which contain elements at infinity.

Thus, in elementary geometry the elements at infinity necessarily play an important role, and differ considerably from ordinary geometric elements in their relationships with the latter. In contrast, in projective geometry, the above-mentioned feature distinguishing elements at infinity from ordinary elements loses its meaning, since metric properties of figures are not its objects. What is more, since elements at infinity may become ordinary elements by projection, they do not possess any projective properties distinguishing them from ordinary elements. Therefore in projective geometry there is no difference between ordinary elements and elements at infinity.

§ 83. The idea of elements at infinity originated a long time ago. But to put the elements at infinity and ordinary elements on equal footing, which is natural from the point of view of projective geometry, remained an illusion up to the time when projective properties of figures were investigated by the methods of elementary geometry, because these methods are based on measurement and the metric of elementary geometry necessarily leads to the difference between the finite images and the images at infinity. In order to precisely define the notion of the projective space, it became necessary to totally omit from projective geometry everything connected with measurement.

The task of freeing projective geometry from the measuring process was accomplished, in principle, by the German mathematician K. G. C. von Staudt (1798-1867).

Projective geometry, freed from the metric, became a discipline studying only properties of the location of geometric figures. At the same time projective geometry grew up as an individual geometrical discipline with its own axioms and its own set of objects (such as the projective line, projective plane and projective space).

5.2 Desargues' theorem. Harmonic sets of elements

§ 84. Projective geometry will be developed on the basis of a system of axioms expressing the mutual relationships between its basic objects. The main objects are points, lines and planes; their mutual relationships mentioned in the axioms are those of incidence and betweenness. Axioms of projective geometry and the theorems obtained from them express the known properties of the Euclidean space augmented by elements at infinity. Of course, any objects may be treated as points, lines and planes of projective geometry and mutual relationships between these objects may be suitably interpreted, the only condition being that the axioms of projective geometry are satisfied. The conclusions thus drawn would represent

definite facts regarding the chosen objects. Accordingly, a projective space is any set of objects, called points, lines and planes; the mutual relationships between these objects satisfy the axioms given below.

The axioms of projective geometry may be divided into three groups:

- Group I contains nine axioms of incidence;
- Group II contains six axioms of betweenness;
- Group III contains one axiom of continuity.

The axioms of Group I and its important consequences will be dealt with in the present section.

Group I. Projective axioms of incidence.

It is assumed that lines and planes have definite correspondence with points which are expressed by the words "a line passes through a point" or "a point lies on a line"; "a plane passes through a point" or "a point lies in a plane". The correspondence must satisfy the conditions of the following Axioms I.1-I.9.

(I.1) *Given any two points A and B , there is a line a passing through A , B .*

(I.2) *Given any two distinct points A and B , there is at most one line passing through A , B .*

(I.3) *There are at least three points on a given line. There are at least three points that do not lie on one line.*

(I.4) *Through any three points A , B , C , not lying on one line, there passes a plane α . Every plane contains at least one point.*

(I.5) *Through any three points A , B , C , not lying on one line there passes at most one plane.*

(I.6) *If two points A , B of a line a lie in the plane α , then every point of a lies in α .*

(I.7) *If two planes α and β have a common point A , then they have at least one more common point B .*

(I.8) *There are at least four points not lying in one plane.*

(I.9) *Two lines in one plane have a common point.*

If Axioms I.1-I.9 just formulated are compared with Hilbert's axioms of Group I (see Chapter 2, § 12), then it may be remarked first that all the requirements of Hilbert's axioms of Group I are also contained in projective Axioms I.1-I.9. Therefore all the theorems of elementary geometry based only on axioms of incidence are also valid in projective geometry. Projective axioms of incidence differ from axioms of incidence of elementary geometry in only two respects.

(1) Axiom I.3 of the projective system requires that a line contain at least three points whereas the corresponding axiom of Hilbert's system requires a line to have at least two points.

(2) Projective axioms of incidence contain Axiom I.9, which is not assumed and not fulfilled in elementary geometry. Because of

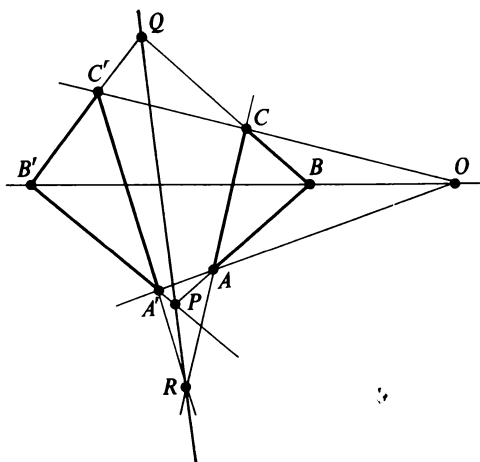


Figure 90

Axiom I.9 there is no parallelism in projective geometry, since any two lines lying in one plane intersect.

Thus the projective axioms of incidence require more conditions to be satisfied than the axioms of incidence of elementary geometry; thanks to this, there are many theorems which can be derived from the projective axioms of incidence but not from Hilbert's axioms of incidence. In particular, Axioms I.1-I.9 imply that

- (i) a line and a plane always have a common point;
- (ii) two planes always have a common line;
- (iii) three planes always have a common point.

§ 85. Omitting the trivial consequences of Axioms I.1-I.9, we proceed directly to the proofs of the Theorems of Desargues, which play a fundamental role in the study of plane projective geometry.

By a *three-point* we shall mean a set of three noncollinear points and three lines joining these points pairwise. The three points in question will be called *vertices* and the lines joining them the *sides* of the three-point (we avoid calling such figures triangles reserving this term for a figure of different form which will be dealt with after discussing the projective axioms of betweenness).

Consider two three-points with vertices A, B, C and A', B', C' . The vertices denoted by the same letter will be called the corresponding vertices (A and A' ; B and B' ; C and C') and the sides containing the corresponding vertices the corresponding sides.

Theorem I (Desargues' first theorem—direct theorem). *If the corresponding sides of two three-points ABC and $A'B'C'$ intersect at points P, Q, R lying on one line, then the lines joining the corresponding vertices are concurrent (Fig. 90).*

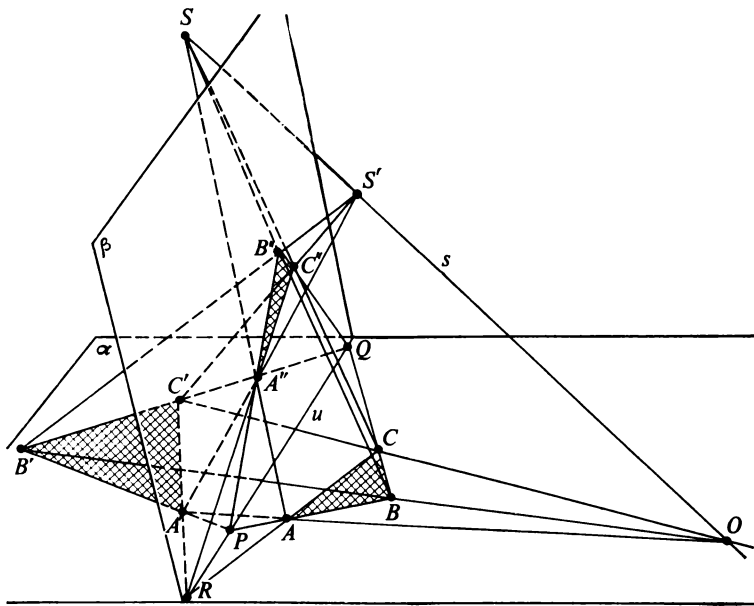


Figure 91

Theorem 2 (Desargues' second theorem—converse theorem). *If the lines joining the corresponding vertices of two three-points ABC and $A'B'C'$ are concurrent, then the corresponding sides of these three-points intersect at points lying on one line.**

The line on which the points of intersection of corresponding sides lie will be called the *axis of perspective* and the point through which the lines joining the corresponding vertices pass will be called the *centre of perspective*. Then the two theorems of Desargues may be formulated in short as follows:

If two three-points have an axis of perspective, then they also have a centre of perspective, and conversely.

We prove Desargues' first theorem.

Let ABC and $A'B'C'$ be two three-points lying in one plane α and u the axis of perspective (Fig. 91). Thus the line u contains points P , Q , R , where the corresponding sides AB and $A'B'$, BC and $B'C'$, AC and $A'C'$ intersect, respectively. We have to show that the lines AA' , BB' and CC' pass through one point, that is, the given three-points have a centre of perspective.**

* Only the case where both the three-points ABC and $A'B'C'$ lie in one plane is important for us.

** It is assumed that line u does not contain the vertices of the given three-points (otherwise the theorem is evidently valid).

For the proof, take a point B'' not lying in the plane α (the existence of B'' is ensured by Axiom I.8). The points P , Q and B'' are non-collinear; therefore there is a unique plane β passing through these points. By Axiom I.3, a point C'' distinct from B'' and Q may be chosen on the line $B''Q$. By Axiom I.6, this point, as well as the point R , lies in β ; accordingly, line RC'' is in β . Since lines RC'' and PB'' lie in one plane, Axiom I.9 implies that they have a common point; call this point A'' . A three-point $A''B''C''$ is thus formed in β and it is located in a special way with respect to the three-points ABC and $A'B'C'$; namely, the three-points ABC , $A'B'C'$ and $A''B''C''$ have a common axis of perspective u , and the corresponding sides AB , $A'B'$, $A''B''$ of the three three-points pass through one point P , while, similarly, the sides BC , $B'C'$, $B''C''$ pass through Q and the sides AC , $A'C'$, $A''C''$ pass through R .

Because of the manner in which the three-points ABC and $A''B''C''$ as well as $A'B'C'$ and $A''B''C''$ are located, they have a centre of perspective. This result may be easily proved. Although this amounts to showing the same fact regarding the three-points ABC and $A''B''C''$ (or regarding $A'B'C'$ and $A''B''C''$) as Desargues' theorem asserts regarding the three-points ABC and $A'B'C'$, the proof here is extremely simplified due to the fact that the three-points ABC and $A''B''C''$ (or $A'B'C'$ and $A''B''C''$) lie in different planes.

Consider the three planes PAA'' , QBB'' and RCC'' . As noted at the end of § 84, any three planes have a common point; denote by S the common point of the three planes in question. Note that the line AA'' is the common line of the planes PAA'' and RCC'' . Next, it is very important to establish that the planes PAA'' and RCC'' are distinct. In fact, the plane PAA'' contains the line BB'' , but, by the choice of B'' , the lines BB'' and u do not have a common point. This means that R does not lie in the plane PAA'' and thus the planes PAA'' and RCC'' are indeed distinct. Thus AA'' , being the common line of the two planes, contains all their common points, including S . In other words, the line AA'' passes through S . Similar arguments show that the lines BB'' and CC'' also pass through S , establishing the existence of the centre of perspective for the three-points ABC and $A''B''C''$. In the same manner, it may be shown that the three-points $A'B'C'$ and $A''B''C''$ have a centre of perspective S' .

Draw a line s through the points S and S' ; this line meets the plane α at some point O . It may be easily established that this point O is the centre of perspective for the three-points ABC and $A'B'C'$. In fact, let us project from centre S the spatial figure formed by the three-points $A'B'C'$ and $A''B''C''$ and the point S' on the plane α . Obviously, $A'B'C'$ is projected onto itself, while the projection of $A''B''C''$ is the three-point ABC . The lines $A'A''$, $B'B''$ and $C'C''$ are projected respectively on the lines AA' , BB' , CC' . And since the lines $A'A''$, $B'B''$ and $C'C''$ pass through S' , their projections,

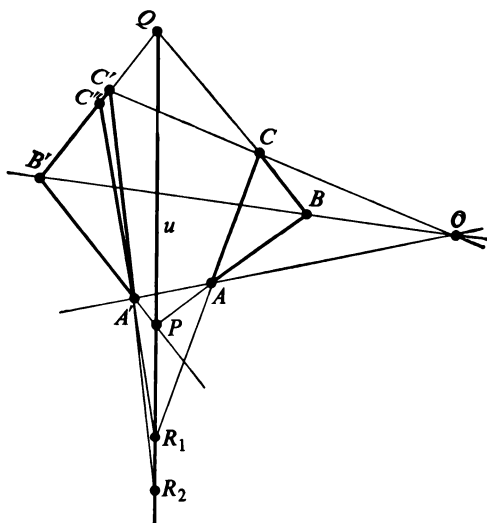


Figure 92

the lines AA' , BB' , CC' , will pass through the projection of S' , that is, through O . Thus it has been shown that the lines joining the corresponding vertices of three-points ABC and $A'B'C'$ pass through one point. ■

We shall now prove the converse theorem.

Let two three-points ABC and $A'B'C'$ be given in a plane which have a centre of perspective, that is, the lines AA' , BB' , CC' pass through one point O . We have to show that these three-points have an axis of perspective, that is, that the points P , Q , R , where the corresponding sides AB and $A'B'$, BC and $B'C'$, AC and $A'C'$ meet, lie on one line.

It will be convenient at the outset to drop from consideration the trivial case when the two three-points have a common side, for example, when the lines BC and $B'C'$ coincide. In this case, the point Q becomes indeterminate and one may assume that this point lies on one line with the points P and R . Thus the theorem is valid in this case. It is assumed below that the corresponding sides of the two three-points ABC and $A'B'C'$ are distinct.

The theorem will be proved by the method of contradiction. Assume that AB and $A'B'$, BC and $B'C'$, AC and $A'C'$ intersect at three points P , Q , R which do not lie on one line. Then the points P and Q are necessarily distinct and determine a line u which intersects the lines AC and $A'C'$ at distinct points R_1 and R_2 such that R_1 , A' and C' do not lie on one line (Fig. 92). Therefore the line R_1A' meets

the line $B'C'$ at a point C'' distinct from C' . The point C'' is not on the line $C'CO$. In fact, if C'' were on the line $C'CO$, then the point B' would also lie on this line and so would point B . But in that case the corresponding sides BC and $B'C'$ must coincide, a case which we have already discarded. Thus the line $C''C$ does not pass through the point O . Consider the three-points ABC and $A'B'C''$. In view of what has just been proved, these three-points do not have a centre of perspective, but they have an axis of perspective, namely u , where the three points P, Q, R_1 lie.

Thus a contradiction has been obtained with Desargues' direct theorem. ■

We now proceed to the most important definition of projective geometry and to the construction of harmonic elements. Our subsequent arguments will be based on Desargues' theorem.

§ 86. A plane figure composed of four points, no three of which are collinear, and six lines joining these points pairwise is called a *complete four-point*.

The points in question are called *vertices* and the lines joining them pairwise are called *sides* of the four-point. In Fig. 93, a four-point has been drawn with vertices A, B, C, D . The sides having no common vertex are called *opposite*. For example, the four-point $ABCD$ has pairs of opposite sides AB and CD , AC and BD , BC and AD . The points of intersection of the opposite sides are known as the *diagonal points* of the four-point. In Fig. 93, the diagonal points are P, Q, R .

With the aid of a complete four-point we define the notion of the harmonic set of elements. A pair of points S, T of any line u is said to be *harmonic conjugate* to the pair of points P, Q of the same line

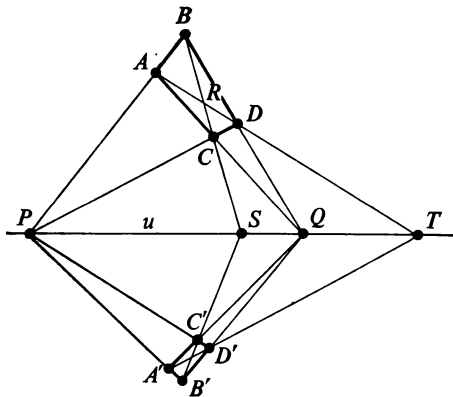


Figure 93

if P and Q are diagonal points of a four-point and S and T are points of intersection of the line u with the pair of opposite sides of the four-point passing through the third diagonal point (Fig. 93).

In this definition, the points P and Q of the first pair enjoy equal status as do the points S and T of the second pair (but so far the same cannot be asserted regarding the pairs P, Q and S, T).

The point T will be called the *fourth harmonic* to the three points P, Q, S if the pair S, T is harmonically conjugate to the pair P, Q (it is important to note that the first two places are occupied by the points constituting the first pair of the harmonic set). Evidently, if T is the fourth harmonic to the three points P, Q, S , then S is the fourth harmonic to the three points P, Q, T .

The stated definition of the harmonic pairs of points contains also the technique for obtaining the fourth harmonic to the three given points. In order to obtain the fourth harmonic to any three given points P, Q, S of a line u , one has to choose in the plane a point B outside u and a point A , distinct from P and B , on the line PB (the existence of A follows from Axiom I.3). Then a point C is determined by the intersection of lines BS and AQ , after which a point D is determined by the intersection of lines PC and BQ . Now by drawing line AD we obtain the desired point T .

The most important thing is to establish the fact that given points P, Q, S , the fourth harmonic point T is located uniquely, that is, its location does not depend on the choice of the points B and A .

This result is a direct consequence of the following

Theorem 3. *Let $ABCD$ and $A'B'C'D'$ be two four-points with common diagonal points P and Q (Fig. 93). If the sides BC and $B'C'$ of these four-points intersect at a point S of the line PQ , then the sides AD and $A'D'$ intersect at a point T of the same line.*

The proof of this theorem is based on Desargues' theorem.

Consider the three-points ABC and $A'B'C'$. Their corresponding sides meet at three points P, S, Q lying on one line. Then Desargues' first theorem implies that the lines AA', BB' and CC' meet at one point O . The corresponding sides of the three-points BCD and $B'C'D'$ also meet at three points of one line, the same points P, Q, S . Another application of Desargues' first theorem shows that the lines BB', CC' and DD' have a common point O' . The points O and O' evidently coincide because both of them are determined by the intersection of lines BB' and CC' . Thus all the lines AA', BB', CC' and DD' intersect at one point O , establishing, in particular, that the lines AA', BB' and DD' meet in one point. Then, by Desargues' second theorem, the corresponding sides of the three-points ABD and $A'B'D'$ intersect at three points of one line. This means that the point T , where the lines AD and $A'D'$ meet, is situated on the line PQ . ■

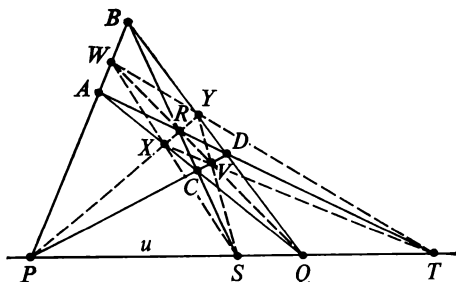


Figure 94

The definition of a harmonic set of points and the theorem just established immediately imply the following proposition expressing the fact that the fourth harmonic point is located uniquely.

Theorem 4. *If the pair S, T is harmonically conjugate to the pair P, Q and $ABCD$ is any four-point with diagonal points P, Q and side BC passing through S , then the side AD passes through T .*

Let us show the following important theorem.

Theorem 5. *If the pair of points S, T of a line u is harmonically conjugate to the pair P, Q , then the pair P, Q is harmonically conjugate to S, T .*

To prove this theorem, take a four-point $ABCD$ having diagonal points P, Q and the pair of opposite sides BC and AD that intersect u at the points S and T (Fig. 94). Denoting by R the third diagonal point of the four-point $ABCD$, draw the lines PR and QR . These lines intersect the sides of the quadrangle at four points X, Y, V, W , as shown in Fig. 94.

Now consider the four-point $AXRW$. This four-point has diagonal points P, Q and its side AR passes through the point T . Since the point S is the fourth harmonic to the three points P, Q, T , it follows from Theorem 4 that the side XW of four-point $AXRW$ must pass through S . Thus the points W, X, S lie on one line. Similar arguments applied to the four-points $RWB Y$, $YDVR$, $RVCX$ show that each of the triples of points W, Y, T ; Y, V, S ; X, V, T lies on one line each.

This implies that $XVYW$ is a four-point with diagonal points S, T and with sides XY, VW passing through the points P, Q , which means that the pair P, Q is harmonically conjugate to the pair S, T . ■

The above theorem establishes the mutual harmonic conjugateness of two pairs of points. Therefore, in what follows, if a pair of points of a line is harmonically conjugate to another pair of points of this line, we will not stress which of the pairs is conjugate to the other pair, and the two pairs will be called *mutually harmonic*.

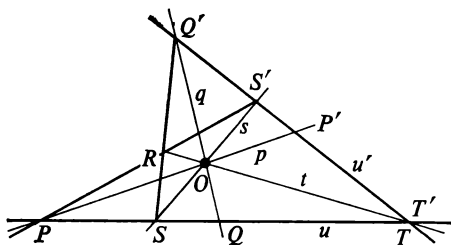


Figure 95

One of the important properties of a harmonic set of points is expressed by

Theorem 6. *Let p, q and s, t be two pairs of rays of a pencil with centre O and let them intersect the line u at pairs of points P, Q and S, T and the line u' at pairs P', Q' and S', T' . Then, if the pairs P, Q and S, T are mutually harmonic, the pairs P', Q' and S', T' are also mutually harmonic.*

Proof. First we consider the particular case where a pair of the corresponding points P, Q, S, T and P', Q', S', T' coincide; for example, the points T and T' coincide (Fig. 95).

Draw the line SQ' and let R be its point of intersection with the line OT . Since the pairs P, Q and S, T are mutually harmonic, the line RS' , by Theorem 4, must pass through P (to see this, consider the four-point $ROS'Q'$ with diagonal points S, T). This results in a four-point $RPSO$ with diagonal points S', T' and with sides OP and RS that pass through the points P' and Q' . This implies that the pairs P', Q' and S', T' are mutually harmonic, proving the theorem for the particular case in question.

Consider now the general case of the location of lines u and u' . Draw the line $P'T$ (Fig. 96). The rays p, q, s, t intersect the line $P'T$ at points P'', Q'', S'', T'' . Since the points T and T' coincide, the fact that the pairs P, Q and S, T are mutually conjugate implies that so are the pairs P'', Q'' and S'', T'' , and, because the points P'' and P' coincide, the harmonic conjugateness of the pairs P'', Q'' and S'', T'' implies the harmonic conjugateness of the pairs $P'Q'$ and S', T' . ■

According to this theorem, if two pairs of rays are met by a line at two pairs of harmonically conjugate points, then the pairs of points obtained by intersecting any other line by the same pairs of rays are also harmonically conjugate. Thus the property that two pairs of rays of a pencil determine harmonically conjugate pairs of points on a given line does not depend on the choice of the line and belongs to the pairs themselves. The pairs of rays possessing this property are known as *harmonically conjugate rays*.

We shall say that the rays $p, q, s, \dots$ originating from O towards $P, Q, S, \dots$ project these points from O . The construction of rays $p, q, s, \dots$ projecting the given points $P, Q, S, \dots$ will be called the *operation of projection*, and the construction of points $P, Q, S, \dots$ on the basis of rays $p, q, s, \dots$ the *operation of section*.

Using this terminology, we can formulate the following statement (as a corollary to Theorem 6):

If the operation of projection or the operation of section are applied to harmonically conjugate pairs of elements (points of a line or rays of a pencil), they result in harmonically conjugate pairs of elements.

Or alternatively,

The property of harmonic conjugateness is invariant under the operations of projection or section.

Using the terminology similar to the one used with respect to a harmonic set of points of a line, we will call the ray t of a pencil the fourth harmonic to the triple of rays p, q, s of the same pencil if the pairs p, q and s, t are harmonically conjugate (in the triple p, q, s first two places are occupied by the pair p, q).

The reader can easily follow the construction of the fourth harmonic ray t on the basis of three given rays p, q, s from Fig. 97 (in Fig. 97, first the lines a and b are arbitrarily drawn through any point of ray p , then line c and, finally, line d are drawn; the position of ray t is then uniquely determined).

§ 87. All the theorems proved in this section concern plane projective geometry. The source of these theorems is Desargues' theorem which, according to its meaning, is also a theorem of plane geometry. However, this theorem was proved by arguments which used space geometry. It is natural to raise the question whether Desargues' theorem can be established without using any spatial configuration.

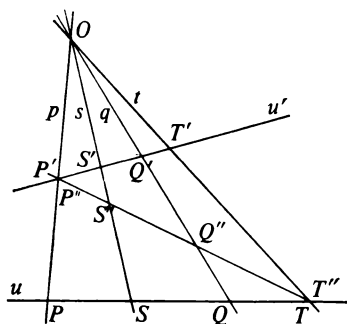


Figure 96

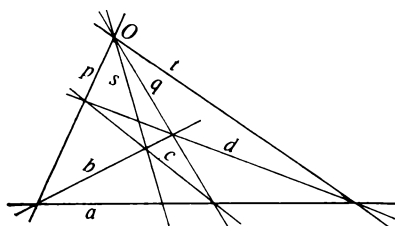


Figure 97

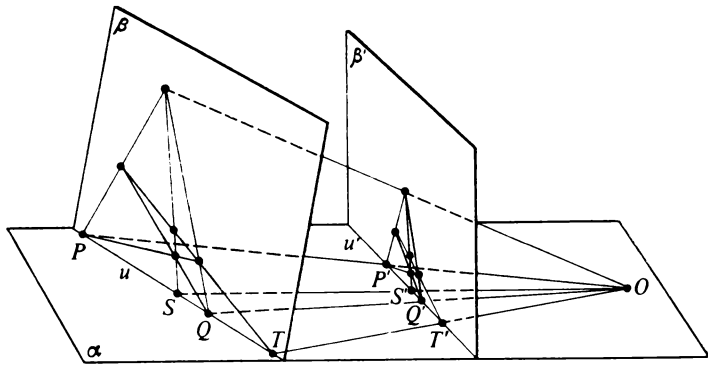


Figure 98

Such theorems exist in literature, but all of them are of a metric nature and hence unacceptable for projective geometry. An analysis of this question undertaken by Hilbert showed that Desargues' theorem cannot be proved within the framework of projective geometry without using spatial configurations.

To be more precise, if all the assertions in the list of Axioms I.1-I.9 that concern space are omitted, then Desargues' theorem cannot be proved by the remaining Axioms I.1-I.3 only. The independence of Desargues' theorem from Axioms I.1-I.3 (even from the axioms of an extended list which includes, apart from Axioms I.1-I.3, projective axioms of betweenness and continuity; these axioms will be discussed later) may be shown in principle by the same method as the one discussed at great length in Chapter 4, where it is used several times. (Hilbert's proof is given in his *Foundations of Geometry*.)

In view of the reasons just mentioned, Desargues' theorem may be treated as an axiom of plane projective geometry.

§ 88. We conclude this section with two remarks. The first remark concerns the proof of the fact that a harmonic set of elements is invariant under the operation of projection. This proposition was proved by using plane geometry. If we were to use spatial configurations, this proposition could be proved in a more obvious manner.

Let u, u' be two lines in a plane α (Fig. 98), P, Q, S, T points of the line u and P', Q', S', T' their projections from centre O on u' (O also lies in α). Assume that the pairs P, Q and S, T are harmonically conjugate. Through u and u' respectively there pass planes β and β' distinct from α . Since the pairs of points P, Q and S, T are harmonically conjugate, a four-point Ω may be constructed in β such that the points P, Q are diagonal points and the pair of opposite sides pass through S and T . Projecting, as on a screen, the four-point

Ω from centre O on plane β' , we obtain a four-point Ω' in β' which bears the same relations with the pairs P', Q' and S', T' as does Ω with the pairs P, Q and S, T . This immediately implies that the pairs P', Q' and S', T' are harmonically conjugate.

The second remark concerns the possible generalization of the theorem on invariance of a harmonic set of elements under projection.

Hitherto projection from a centre was discussed. But sometimes, besides central projection, one has to consider (in space projective geometry) projection from an axis.

Let o be a line and $P, Q, S, \dots$ be a system of points on another line u which does not lie in one plane with o (Fig. 99). A system of planes $\pi, \kappa, \sigma, \dots$ passing through line o and points $P, Q, S, \dots$ is called the pencil of planes with axis o projecting the points $P, Q, S, \dots$. If u' is any other line intersecting the planes $\pi, \kappa, \sigma, \dots$ at the points $P', Q', S', \dots$, then P', Q', S' are said to have been obtained by axial projection of $P, Q, S, \dots$.

It turns out that the invariance of harmonic conjugacy of pairs of points is preserved under an axial projection.

Let an axial projection map a harmonic quadruple of points P, Q, S, T of a line u into a quadruple of points P', Q', S', T' of another line u' (in general, the lines u and u' do not lie in one plane). We shall show that the quadruple P', Q', S', T' is also harmonic.

To do this, draw line u'' intersecting both u and u' (Fig. 99). Lines u and u'' lie in one plane α and lines u' and u'' lie in one plane β . Let A, B be the points where planes α, β meet the axis of the projecting pencil of planes and P'', Q'', S'', T'' be the points where the planes of the pencil meet line u'' . Evidently, points P'', Q'', S'', T''

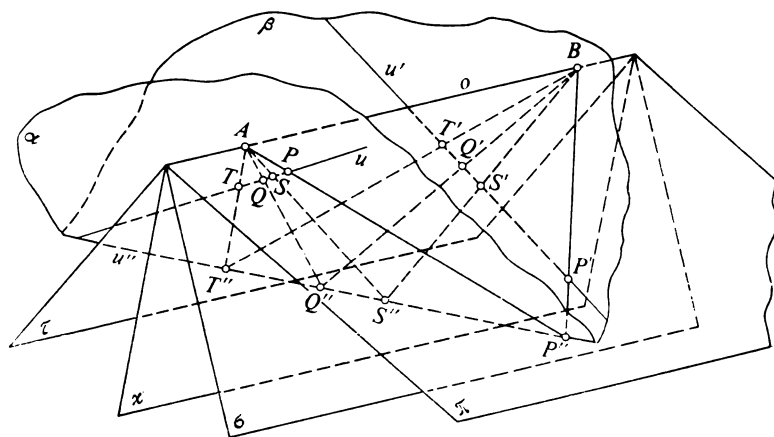


Figure 99

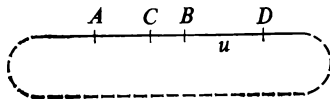


Figure 100

may be treated as having been obtained by projecting P, Q, S, T from the centre A inside the plane α . Therefore the fact that the set of points P, Q, S, T is harmonic implies that the set of points P'', Q'', S'', T'' is also harmonic. Next, treating the points P', Q', S', T' as having been obtained by central projection of P'', Q'', S'', T'' from centre B , we may conclude, noting that the quadruple P'', Q'', S'', T'' is harmonic, that the quadruple P', Q', S', T' is also harmonic.

Thus, if by intersecting a line two pairs of planes π, κ and σ, τ of a certain pencil determine two pairs of harmonically conjugate points, then they determine two pairs of harmonically conjugate points by intersecting any other line. In this case, the pairs of planes π, κ and σ, τ are called *harmonically conjugate*.

It is easy to see that if a plane not passing through the axis of a pencil intersects the pairs of harmonically conjugate planes of this pencil, then in this intersecting plane two pairs of harmonically conjugate rays of a linear pencil are obtained. The proof of this fact is straightforward and is omitted.

5.3. Order of points on the projective line

§ 89. As we know, the definition of the order of points on a line in elementary geometry is based on the notion of a point lying between two other points (see Chapter 2, § 13). In projective geometry, where a line is treated as a closed curve, this notion becomes meaningless because in considering any three points of a projective line (or three points of a circle) it is not possible to label any particular point as lying between the other two.

To be able to define order among points on a projective line, one has to start with two pairs of points. We shall first use a diagram as a visual aid. Let A, B, C, D be four points of a projective line u located as shown in Fig. 100 (here the projective line has the form of a closed curve). If we want to displace C along u so as to make it coincide with D , then at least for an instant C has to coincide with either A or B . Similarly, for A to coincide with B , the point A must pass through either C or D . In this case, the pair A, B is said to divide the pair C, D .

In the same set of points A, B, C, D , the pairs A, D and B, C are such that there is no need for B or C to pass through A or D if B and C are to coincide or, similarly, if A and D are to coincide

there is no need for A or D to pass through B or C . Regarding the pairs A, D and B, C we say that they do not divide each other. Similarly, the pairs A, C and B, D do not divide each other. In this manner, our visualization of a projective line (or a circle) enables us to distinguish between dividing and nondividing pairs of points.

In the logical development of projective geometry, the division of pairs of points is taken as the basic "betweenness" relation. The necessary properties of this relation are set out in the axioms of the second group.

Group II. Projective axioms of betweenness.

It is assumed that two points of a line have a certain relationship with two other points of this line, and this relationship is expressed by the word "divide". This relationship satisfies the following axioms which are referred to as the axioms of betweenness.

(II.1) *Given any three distinct points A, B, C of a line u , there is a point D on this line such that the pair A, B divides the pair C, D . If the pair A, B divides the pair C, D , then all the four points A, B, C, D are distinct.*

(II.2) *If the pair A, B divides the pair C, D , then the pair B, A divides the pair C, D and the pair C, D divides the pair A, B (that is, the division property is mutual and does not depend on the order of the points of the pair).*

(II.3) *Given any four distinct points A, B, C, D of a line u , it is always possible to form uniquely two dividing pairs.*

(II.4) *Given points A, B, C, D, E on a line u , if the pairs C, D and C, E divide the pair A, B , then the pair D, E does not divide the pair A, B (Fig. 101).*

(II.5) *Given points A, B, C, D, E on a line u , if the pairs C, D and C, E do not divide the pair A, B , then the pair D, E does not divide the pair A, B (Fig. 102).*

(II.6) *Let A, B and C, D be two pairs of points of a line u and let A', B' and C', D' be their projections from a centre to a line u' . If the pairs A, B and C, D divide each other, then so do the pairs A', B' and C', D' . In short, the division property of two pairs of points is invariant under projection.*

On the basis of Axiom II.6, divisions of pairs of lines of a plane pencil may be defined.

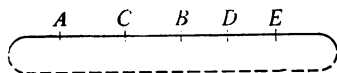


Figure 101

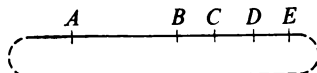


Figure 102

Namely, if a, b and c, d are pairs of lines passing through a point and s is any line intersecting a, b and c, d at the points A, B and C, D , respectively, then Axiom II.6 implies that for any choice of s the pairs A, B and C, D always either divide or do not divide each other. In the former case, we shall say that the pairs of lines a, b and c, d divide each other, while in the latter case the pairs of lines do not divide each other. Thus the notion of division of pairs of lines reduces to that of pairs of points, but the latter is a fundamental notion for us and cannot be replaced by a more primitive notion.

In discussing projective geometry, we shall not attempt to develop it on the basis of a minimum number of requirements. Therefore we shall not examine the question as to whether all the axioms formulated are indeed necessary or whether some of these may be proved as theorems. It is important that these axioms suffice for proving theorems constituting the subject matter of projective geometry.*

Theorem 7. *Let A, B be two given points of a line u . Then all the points of u may be divided into two classes so that any two points of the same class form a pair not dividing A, B , while any pair of points from different classes divides the pair A, B .*

Proof. By Axiom I.3, there is a point C , distinct from A, B , on u . To the first class we assign C and also those points of u which form with C a pair not dividing the pair A, B . To the second class is assigned every point which forms with C a pair dividing the pair A, B . Then all the points of u (except A and B) are classified into two classes. We have to show that this classification satisfies the conditions stated in the theorem.

Take two points C_1 and C_2 of the first class. According to the construction, the pairs C, C_1 and C, C_2 do not divide the pair A, B . Therefore, by Axiom II.5, the pair C_1, C_2 does not divide the pair A, B . Further, if D_1 and D_2 are two points of the second class, then, by construction, C, D_1 and C, D_2 divide A, B . Consequently, by Axiom II.4, the pair D_1, D_2 , as in the first case, does not divide A, B . Thus, if two points lie in the same class, they do not divide A, B .

Take now two points M and N from two different classes. For example, suppose M is in the first class and N in the second. Then the pair C, M does not divide the pair A, B but the pair C, N divides the pair A, B . If the pair M, N does not divide the pair A, B , then, since in this case the pair M, C does not divide the pair A, B , Axiom II.5 implies that the pair C, N does not divide the pair A, B , contradicting the assumption made. Accordingly, M, N divide A, B . ■

It is worth noting that if the above two classes are constructed not by means of C but by using other point of the first class, then

* These axioms together with the axiom of continuity, presented in § 94, constitute a complete system.

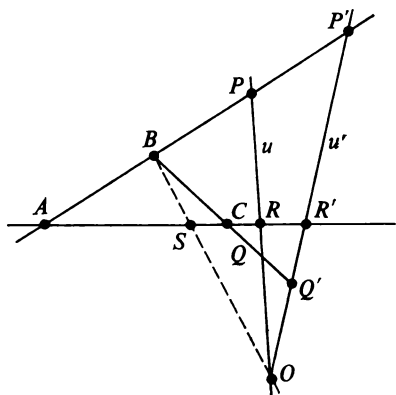


Figure 103

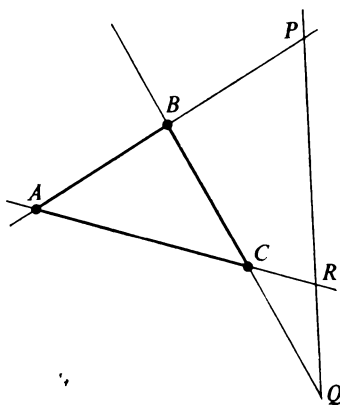


Figure 104

the same two classes are obtained. And if the construction is carried out by means of a point in the second class, then again the old classes are obtained but in reverse order.

Using the common terminology of visual geometry, each of the two classes occurring in the above construction will be called a *segment*. Then the above theorem may be restated in the following form.

Two points A, B of a line divide the points of this line into two classes; if the points M and N lie on the same segment, then the pair M, N does not divide the pair A, B and if the points M and N lie on different segments, then the pairs M, N and A, B divide each other.

In order to distinguish between the two segments in question, a certain point of this segment must be mentioned. Therefore, in projective geometry, a segment is sometimes denoted by three letters; for instance, ACB denotes a segment with endpoints A, B and an interior point C . If the pair C, D divides A, B , then ACB and ADB are two distinct segments with endpoints A, B . The segments ACB and ADB are called *complementary* to each other.

We shall now prove a theorem that will enable us to define a figure in projective geometry that is completely analogous to the Euclidean triangle.

Theorem 8. *Let A, B, C be three noncollinear points and u, u' two lines not passing through either A, B or C . Assume further that P, Q, R are points at which u intersects the lines AB, BC, AC and P', Q', R' are points where u' intersects the same lines. Then, provided the pair P, P' does not divide the pair A, B and the pair Q, Q' does not divide the pair B, C , the pair R, R' does not divide the pair A, C (Fig. 103).*

Proof. Let O be the point of intersection of the lines u and u' . Projecting the pairs A, B and P, P' from O as the centre on line AC ,

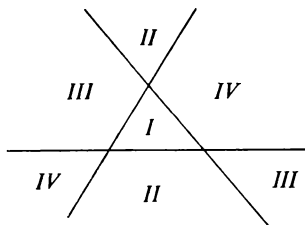


Figure 105

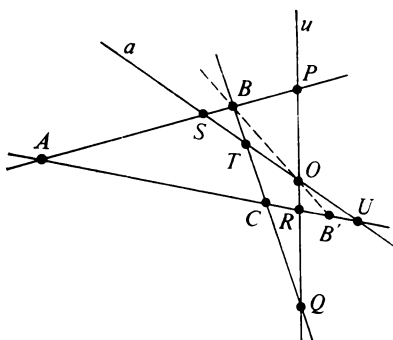


Figure 106

we obtain the pairs A, S and R, R' . By hypothesis, the pairs A, B and P, P' do not divide each other. Then, by Axiom II.6, the pairs A, S and R, R' do not divide each other either. Projecting from centre O the pairs B, C and Q, Q' on line AC , we obtain the pairs S, C and R, R' . Since the pair B, C does not divide the pair Q, Q' , it follows, by the same Axiom II.6, that S, C and R, R' do not divide each other. Thus, the pairs S, A and S, C do not divide the pair R, R' . Therefore Axiom II.5 implies that the pair A, C does not divide R, R' . ■

Taking three noncollinear points A, B, C , choose one of the two segments with endpoints A, B and one of the two segments with endpoints B, C (in Fig. 104, the chosen segments are shown by heavy lines). Let AB and BC denote the segments thus chosen. Choose a point P on the segment complementary to AB and a point Q on the segment complementary to BC and draw the line PQ . Let R be the point of intersection of the lines PQ and AC . Now we arbitrarily change the positions of P and Q , retaining them, however, on the segments complementary to AB and BC , respectively. Then the previous theorem immediately implies that the point R , while displaced along the line AC , always remains inside one of the two segments determined by the points A and C . The segment with the endpoints A and C that is complementary to the segment containing R will be denoted by AC . We see that the segment AC is uniquely determined by the segments AB and BC . The figure formed by the points A, B, C , and the segments AB, BC and AC will be called a *triangle* and the segments AB, BC and AC its *sides*.

It may be easily shown that every three-point ABC determines four triangles with common vertices A, B, C . The sides of these triangles are mutually complementary segments on lines serving as the sides of the three-point. Figure 105 shows triangles I, II, III, IV determined by one three-point ABC .

We shall now show that Pasch's axiom (see Chapter 2, § 13) holds in projective geometry, that is, if a triangle ABC and a line a passing neither through A , B nor C in its plane are given and if this line passes through a point of side AB , then it passes through either a point of side BC or a point of side AC .

To prove this, we note first that, by the definition of a triangle, there is a line u which intersects the lines AB , BC , AC at points P , Q , R , respectively, such that point P lies on the segment complementary to AB , point Q lies on the segment complementary to BC and point R on the segment complementary to AC (Fig. 106). Next, since we are discussing projective geometry, Axiom I.9 implies that the given line a has a common point T with the line BC and a common point U with the line AC . Let S be the point of intersection of a and AB .

Assume that T lies on the segment BQC and U on the segment ARC . Then, by Theorem 8, point S must lie on the segment APB , which contradicts the condition that S lies on the segment AB . Thus line a intersects either the side BC or the side AC of the given triangle. This proves Pasch's axiom. ■

§ 90. In the projective space, take a plane α_∞ . We shall call it the "plane at infinity". All the points and lines lying in α_∞ will be assumed to be "at infinity". The remaining elements of the space will be referred to as "finite". (The terms "finite" and "at infinity" are in quotation marks because the plane α_∞ has been chosen arbitrarily and the difference between "finite" elements and elements "at infinity" is conditional.)

Obviously, every "finite" line contains only one point at infinity, namely the point of intersection of this line with α_∞ . In the set of remaining, that is, "finite", points of any "finite" line, we introduce the "betweenness" relation by means of a completely defined condition which is common to all the lines.

Let a be any "finite" line and O_∞ its "point at infinity". Take any three "finite" points A , B , C of a . If the point B forms with O_∞ a pair B, O_∞ that divides the pair A, C , then we shall say that in the set of "finite" points of a the point B lies between the points A, C . It may be easily shown that the "betweenness" notion thus defined satisfies Hilbert's axioms of betweenness II.1-II.3.

Indeed, according to projective Axiom II.2, if the pair B, O_∞ divides the pair A, C , then it also divides the pair C, A . Consequently, if, by our definition, B lies between A and C , it also lies between C and A . This shows that Hilbert's Axiom II.1 is valid.

Furthermore, whatever the "finite" points A and C , there is always, by projective Axiom II.1, a point D , such that the pair C, O_∞ divides the pair A, D . Accordingly, in the set of "finite" points of a there is always a point D such that C lies between A and D . Thus, Hilbert's Axiom II.2 is also satisfied.

Finally, by projective Axiom II.3, from the four points A, B, C, O_∞ two dividing pairs can be formed in only one way. Thus, of three points A, B, C only one lies between the other two. Accordingly, Hilbert's Axiom II.3 is satisfied.

Having established the "betweenness" notion in the set of "finite" points of a line, one may state the definition of a segment, the same as in visual geometry, as a collection of points of a line which lie between any two of the line's points. Obviously, a segment thus defined is nothing but one of the two mutually complementary segments in which the points A, B divide the projective line passing through these points, namely, the segment which does not contain the point at infinity.

It is also obvious that a triangle in the sense of the relationships introduced in the system of "finite" elements of the projective space is simultaneously a triangle in the sense defined in projective geometry (see § 89). Therefore one may assert that Pasch's axiom is valid for the system of "finite" elements of the projective space because this axiom has been shown to be valid for the projective space as a whole.

Thus the "betweenness" notion has been defined for the system of "finite" elements so that all the Hilbert's axioms of order are true.

Now suppose that the plane "at infinity" α_∞ is completely omitted, together with points and lines "at infinity" contained in this plane, from the projective space, or, to use a common phrase, the projective space is cut along the plane α_∞ . It is easily seen that the remaining elements satisfy Hilbert's axioms of order. Thus, on the basis of this fact and the previous result, we can assert that all theorems of elementary geometry based only on Hilbert's axioms of the first two groups are valid for the projective space cut along one of its planes. In particular, one can assert that there are an infinite number of points, lines and planes in the projective space.

The process just described is the opposite of the one in the preceding section. It was shown there that by introducing new elements into it, a Euclidean space may be turned into a projective space. Here we find that if a certain plane is deleted from the projective space defined by means of special axioms, it turns into a space analogous to the Euclidean space.

We shall use, the method of cutting the space along one of the planes to prove a number of projective theorems; in the process we shall use the results that follow from the first two groups of axioms of elementary geometry, which were studied in Chapter 2. In particular, we shall now use this method to establish the order of points characteristic of a projective line.

Let A and B be two points of a projective line u ; these points divide u into two mutually complementary segments. Consider one of them, calling it AB . We introduce an order for the set of interior

points of AB by assuming that M precedes N if the pair A, N divides the pair M, B . We have to show that this order relationship is transitive, that is, if M precedes N and N precedes P , then M precedes P . To do this in the simplest possible way, we introduce a point at infinity on u . For this point at infinity we take B . Then the condition " N precedes P " may be expressed thus: "in the set of finite points of u the point N lies between A and P " or " N lies inside the segment AP ". In this case, for proving the transitivity of the order relation introduced, it suffices to prove the following proposition for the set of finite points of u : "if M lies inside the segment AN and N lies inside AP , then M also lies inside AP ". But this result was obtained earlier from the first two groups of Hilbert's axioms (see Chapter 2, § 14).

We shall say that the order defined on AB corresponds to the direction of the segment from A to B . On the segment complementary to AB , we introduce an order corresponding to the direction from B to A . Having done this, a completely defined order relation may be introduced inside any segment ST of line u by requiring that for the common portions of the segment ST with the segment AB and its complement the order of points of ST coincides with the order of points of these segments. Considering all the possible cases of location of ST , namely (i) the segment ST is contained in the segment AB , (ii) the segment ST is contained in the segment complementary to AB , (iii) the segment ST covers the segment AB , (iv) the segment ST covers the segment complementary to AB , (v) the segment ST covers a part of segment AB and a part of the complementary segment, the reader will easily find that in all these cases the set of interior points of the segment ST may be uniquely ordered and this order satisfies the above-mentioned conditions.

The property of mutual location of points of a projective line due to which on every segment an order relation satisfying the above-mentioned conditions is induced will be called a *cyclic order*. Depending on the way the initial segment AB has been ordered—from A to B or from B to A —two distinct cyclic orders are established on a projective line. They happen to be the reverse of each other in the sense that if, according to the first of the two orders, among the points M, N lying inside the segment ST point M precedes N , then in the second cyclic order for the points of ST point M follows N .

Axioms II.1-II.6 are referred to as projective axioms of betweenness, since they justify the cyclic order introduced on a projective line.

§ 91. We close the present section by introducing a topology on the projective line, that is, we define the closeness of its points. This can be achieved by constructing a system of neighbourhoods for every point of the projective line.

Consider a projective line u . By a neighbourhood of a point M of u we mean an open segment (that is, a segment with endpoints excluded) containing M . The following propositions hold (on them the topological theorems of elementary analysis are based).

- (1) Each neighbourhood of M contains this point.
- (2) The common part of two neighbourhoods of M contains a neighbourhood of M .
- (3) A neighbourhood of M is also a neighbourhood of any other point of this neighbourhood.
- (4) For two distinct points M and N there are neighbourhoods without common points.

Of these assertions, the first and the third are direct consequences of the definition of a neighbourhood; the second and fourth, although visually obvious, require proofs.

In order to simplify the proofs as far as possible, the projective line should be cut and thus everything reduced to the consideration of segments in the Euclidean sense. The detailed proofs are, however, omitted.

By constructing a system of neighbourhoods on the projective line, we have opened up the possibility of speaking about limiting points of sets, limits of point sequences, continuity of functions defined on the projective line, etc.; generally, about many facts that are termed topological. This possibility will be exploited in the following sections.

5.4. Division of harmonic pairs.

Continuity of the harmonic correspondence

§ 92. It is important for the sequel to show that the diagonal points of a complete four-point are noncollinear.

Let $ABCD$ be a complete four-point with the diagonal points P, Q, R (see Fig. 93). We have to show that line PQ does not pass through R . We cut the plane along the line PQ (that is, the line PQ is deleted), and for the remaining elements we introduce order as in § 90. Then the axioms of Groups I and II of elementary geometry are valid.

In the sense of the order relation defined, the point D lies on the same side of line AB as the point C and on the same side of line AC as the point B .

Consequently, point D lies inside $\angle BAC$. On the basis of this result and Theorem 11a, § 16 of Chapter 2, it may be concluded that the line AD intersects the line BC , that is, these lines have a common point. This means that point R has not been removed in the cut performed, proving the assertion. From this we have the following

Corollary. *If P, Q, S are three distinct points of a line and T is the fourth harmonic to them, then all the four points P, Q, S, T are distinct.*

From Fig. 94, where the construction of mutually harmonic pairs P, Q and S, T is illustrated, it is easy to obtain the following visually obvious theorem.

Theorem 9. *Mutually harmonic pairs of points divide each other.*

This theorem will play an extremely important role later. We present its rigorous proof.

Let two pairs of mutually harmonic points P, Q and S, T be given on a line u . Consider a quadrangle $ABCD$ with P, Q as diagonal points and S, T lying on the opposite sides BC and AD that pass through the third diagonal point R (Fig. 94).

From centre B project P, Q, S, T on line AD ; their projections are A, D, R, T , respectively. The points thus obtained are now projected on u from centre C . The projections of A, D, R, T are Q, P, S, T , respectively.

Thus after two projections the set $PQST$ is mapped into itself but its individual members change their positions as shown in $\begin{pmatrix} PQST \\ QPST \end{pmatrix}$, where the lower row represents the points into which the points of the upper row are mapped by the above process.

Note that the four points P, Q, S, T may be classified into two pairs in only three ways: (a) $(PQ), (ST)$; (b) $(PS), (QT)$; (c) $(PT), (QS)$. It may be easily seen that the pairs (PS) and (QT) do not divide each other. In fact, if they did, then, by Axiom II.6, (QS) and (PT) would also be dividing pairs since they are obtained by projecting the pairs (PS) and (QT) twice. Hence, in this case, the second and third of the three possible classifications of the points $PQST$ yield pairs which divide each other. But this contradicts Axiom II.3, according to which dividing pairs of four points can be formed in only one way.

In exactly the same way, if we assume that (PT) and (QS) are dividing pairs, then again we will arrive at a conclusion that the pairs (PS) and (QT) divide each other, which is again a contradiction. Since, by Axiom II.3, one of the three ways of forming pairs out of four points necessarily leads to dividing pairs, it follows that the dividing pairs are the pairs (PQ) and (ST) . ■

This theorem proves more useful in the following formulation. If the pair M, M' is harmonically conjugate to the pair A, B , then the points M, M' lie on different complementary segments determined by the points A, B . If the points A, B are fixed, then M' depends only on M ; this fact is symbolically expressed by writing $M' = f(M)$. Since $MM'AB$ and $M'MAB$ are in equal measure harmonically situated sets of points, the relation $M = f(M')$, besides the relation $M' = f(M)$, also holds. The correspondence

$M' = f(M)$ is called harmonic. Obviously, mutually complementary segments with common endpoints A, B are mapped into each other in a one-to-one manner by a harmonic correspondence. This mapping will be examined below in detail.

§ 93. Take three points M_1, M_2, M_3 on a projective line u . Let M be the fourth harmonic to the three chosen points, namely M is a point such that the pair M, M_3 is harmonically conjugate to the pair M_1, M_2 . We shall use the notation $M = f(M_1, M_2, M_3)$, treating M as a function of three points M_1, M_2, M_3 . Obviously, $f(M_1, M_2, M_3) = f(M_2, M_1, M_3)$, and, if $M = f(M_1, M_2, M_3)$, then $M_3 = f(M_1, M_2, M)$.

If we fix M_1 and M_2 and write $M_1 = A, M_2 = B$ and M instead of M_3 , then the function $f(A, B, M)$ coincides with the function $f(M)$ which was examined at the end of the preceding subsection.

The following important theorem holds.

Theorem 10. *The function $M = f(M_1, M_2, M_3)$ is continuous for all possible locations of points M_1, M_2, M_3 .*

In the light of the definition of neighbourhoods of points of a projective line given in § 91, this theorem may be formulated as follows: Whatever the points M_1, M_2, M_3 and whatever the open segment Δ containing point M , three open segments $\Delta_1, \Delta_2, \Delta_3$ containing points M_1, M_2, M_3 , respectively, may always be found such that if these points remain inside, while varying, the segments $\Delta_1, \Delta_2, \Delta_3$, then point M remains inside, while varying, segment Δ .

Thus the proof of Theorem 10 must be purely constructive because it leads to the construction of segments $\Delta_1, \Delta_2, \Delta_3$ on the basis of segment Δ . We shall carry out the proof for two particular cases only, which are of interest to us: (a) the third argument of function $f(M_1, M_2, M_3)$ is a fixed point $M_3 = C$, and (b) the first two arguments of $f(M_1, M_2, M_3)$ are fixed points $M_1 = A, M_2 = B$.

In the first case we have a function of two arguments, $M = f(M_1, M_2, C)$. Suppose that definite positions of M_1 and M_2 are given. The corresponding point M is illustrated in Fig. 107a, where Q_1, Q_2, G, H is a four-point with diagonal points M_1, M_2, Q and one side Q_1Q_2 passing through C and the other side HG through M .

Assume that the projective plane is cut along the line Q_1Q_2 . Then in discussing a certain segment with known endpoints we need not point out which one of the two mutually complementary segments with given endpoints is being considered because in the projective plane thus cut any two points determine a segment uniquely.

Figure 107b illustrates the same construction with the same notations, but now the line Q_1Q_2 is visually removed to infinity; in this figure the lines going towards Q_1 (or towards Q_2 or Q) are parallel and M is the midpoint of the segment M_1M_2 . For further discussion, it is easier to follow Fig. 107b.

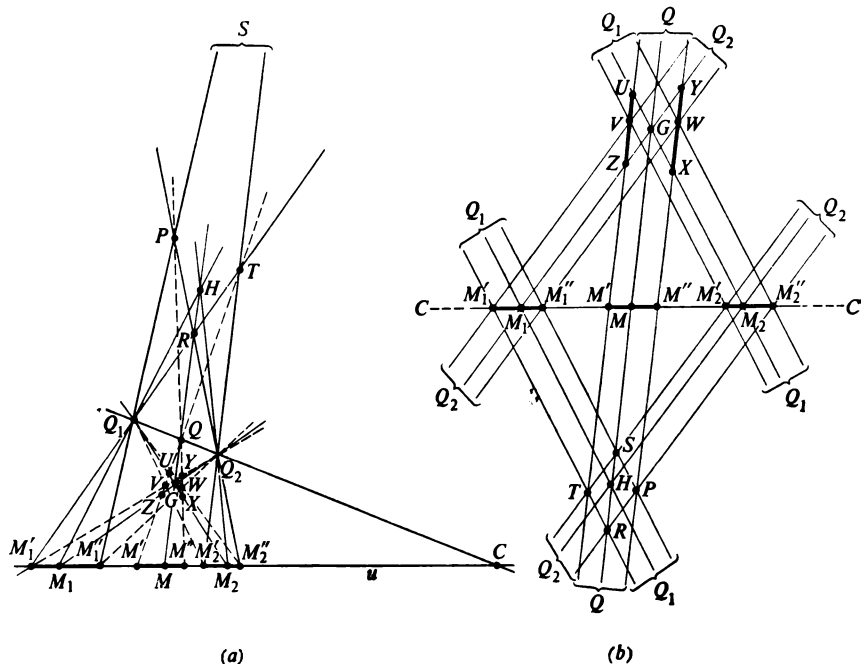


Figure 107

Suppose that M_1 and M_2 change their positions on line u . Since C remains unchanged, for finding M we can always use the four-point with fixed vertices Q_1 and Q_2 . Then, as can be easily verified, the diagonal point Q also remains fixed since it is the fourth harmonic to the three fixed points Q_1, Q_2, C .

If the point M_1 remains, while varying, inside a segment $M_1'M_1''$ and M_2 remains, while varying independently of M_1 , inside a segment $M_1'M_1''$, then the variable vertex H of the four-point, with whose aid the point M is found on the basis of points M_1, M_2, C , remains, while varying, inside the quadrangle $PSTR$. The projection of H from Q to u yields just M . When the points M_1 and M_2 occupy the extreme positions M_1' and M_2' , point H coincides with T and M falls upon a point M' ; when M_1 and M_2 occupy the extreme positions M_1'' and M_2'' , point H coincides with P , and M falls upon M'' . For the remaining positions of M_1 and M_2 the point M remains between M' and M'' .

It may now be easily conceived how, by preassigning a neighbourhood $\Delta = M'M''$ of point M , to construct neighbourhoods Δ_1 and Δ_2 of points M_1 and M_2 such that if these points remain, while varying, inside Δ_1 and Δ_2 , respectively, then M remains,

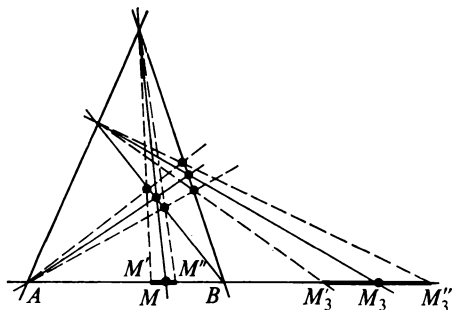


Figure 108

while varying, inside Δ . To do this, one has to proceed in the following manner: project M' and M'' from Q ; on the projecting rays mark off segments ZU and XY on the lines Q_1M_2 and Q_2M_1 ; choose any point V inside ZU and any point W inside XY . Projections of V, W from Q_2 to u will be the points M'_1, M''_1 , which close the segment $\Delta_1 = M'_1M''_1$; projections of the same points V, W from Q_1 will be M'_2, M''_2 , which close the segment $\Delta_2 = M'_2M''_2$.

The segments Δ_1 and Δ_2 thus constructed are the desired neighbourhoods of M_1 and M_2 , that is, if M_1 and M_2 change their positions but remain inside Δ_1 and Δ_2 , respectively, then $M = f(M_1, M_2, C)$ changes its position but remains inside Δ . Since the neighbourhoods Δ_1, Δ_2 with such properties may be constructed for any given points M_1, M_2 and neighbourhood Δ , it follows that $M = f(M_1, M_2, C)$ is continuous for any positions of points M_1 and M_2 on the projective line u .

Turning to the second aforementioned case, namely when $M_1 = A$ and $M_2 = B$ are fixed in the relation $M = f(M_1, M_2, M_3)$, we refer to Fig. 108; this figure makes clear, without any explanation, how, by preassigning a neighbourhood $\Delta = M'M''$ of M , to construct neighbourhood $\Delta_3 = M'_3M''_3$ of M_3 such that when M_3 varies remaining inside Δ_3 the point M remains inside Δ . That such a construction is possible implies that the function $M = f(A, B, M_3)$ is continuous.

We shall not prove the continuity of the function $M = f(M_1, M_2, M_3)$ in all its arguments since for our purpose Theorem 10 in its general form is not required. On the other hand, the function of one argument, $M' = f(A, B, M)$, will be examined from another point of view.

Suppose that a cyclic order is introduced on line u such that the set of points of one of the segments with endpoints A, B is ordered in the direction from A to B and the set of points of the complementary segment is ordered in the direction from B to A . Let AB denote the first of these segments.

If M lies inside the segment AB , then, by Theorem 9, $M' = f(A, B, M)$ must lie on the segment complementary to AB . Let M_1 and M_2 be two arbitrary positions of M inside AB and let M'_1, M'_2 be the corresponding positions of M' . We shall now show that as soon as M_1 precedes M_2 on AB , point M'_1 follows M'_2 on the complementary segment (Fig. 109).

For the proof, construct the complete four-points CDP_1Q_1 and CDP_2Q_2 with common diagonal points A, B such that the opposite sides DP_1 and CQ_1 of the first quadrangle pass through M_1 and M'_1 and the opposite sides DP_2 and CQ_2 of the second through M_2 and M'_2 . From centre D project the set of points ABM_1M_2 on line CB . This yields a set of points CBP_1P_2 . This set is projected from centre A to line DB . The set of points CBP_1P_2 is then mapped into DBQ_1Q_2 . Finally, projecting the set of points DBQ_1Q_2 from centre C to line u , we obtain the set of points $ABM'_1M'_2$. Thus a chain of projections maps the points ABM_1M_2 into $ABM'_1M'_2$, respectively.

If M_1 precedes M_2 in the order introduced on the segment AB , then, according to the definition of this order relation, the pair A, M_2 divides the pair M_1, B . In view of the invariance of the division of two pairs under projections (see Axiom II.6), the pair B, M'_1 must divide the pair A, M'_2 . And this means that M'_1 follows M'_2 if the set of points of the complementary segment to AB is ordered in the direction from B to A .

The result obtained can be described clearly in the following way: When a variable point M moves along the segment AB in the direction from A to B , its harmonic conjugate M' moves in the opposite direction along the segment complementary to AB , that is, also from A to B .

If a set of points $M_1, M_2, \dots, M_n$ are located on the segment AB so that the point with the smaller index precedes each point with the larger index, then with the given set there are associated points $M'_1, M'_2, \dots, M'_n$ on the segment complementary to AB such

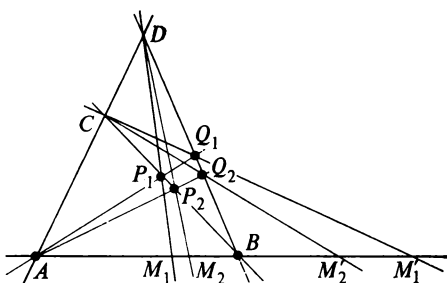


Figure 109

that every point with smaller index follows each point with larger index.

A mapping of a directed segment into another (directed) segment under which the order of points of any ordered set is either always preserved or always goes into reverse order is called an *ordering* (*direct* or *reverse*, respectively) *mapping*. Employing this terminology and taking note of what has been described until now, we may formulate the following

Theorem 11. *The harmonic mapping $M' = f(A, B, M)$ of the segment AB into its complementary segment is a continuous and reverse ordering mapping.*

Remark. So far the points of the harmonic set were supposed to be distinct, and the definition of the fourth harmonic to the given three points concerned only this case. Therefore, for the case where the arguments M_1, M_2, M_3 in the function $M = f(M_1, M_2, M_3)$ coincide, Theorem 10 must be, strictly speaking, proved separately. Without discussing this question for the function $M = f(M_1, M_2, M_3)$, we examine the function $M' = f(A, B, M)$ with $M = A$ and $M = B$.

Let AB be one of the two segments determined on the projective line by the points A and B . Theorem 11 implies that if M moves closer, remaining inside AB , to A monotonically, then M' moves closer, remaining inside the complement of AB , towards A for meeting M . Therefore, if we wish to define the function $M' = f(A, B, M)$ with $M = A$ so that it becomes continuous for this position of M , we must take, apart from $M = A$, also $M' = A$, that is, we must assume $f(A, B, A) = A$. Similarly, $f(A, B, B) = B$.

In the mapping of the projective line into itself in which point $M' = f(A, B, M)$ corresponds to point M , and points A and B correspond to themselves. They are known as *double* (or *fixed*) *points* of the harmonic mapping.

5.5. The axiom of continuity. Projective system of coordinates on a line

§ 94. We have now reached a very important stage in the exposition of projective geometry—the enunciation of the principle that determines a point of projective space by means of coordinates.

In Euclidean geometry the coordinates of points are very easily determined with the aid of measurement. In projective geometry, where the axioms of congruence are absent, the construction of a system of coordinates requires some ingenuity. In the exposition of this question we follow F. Klein.

Besides the two groups of projective axioms—axioms of incidence and betweenness—we shall also require Dedekind's axiom of con-

tinuity, the only axiom of the third group. For its formulation we introduce necessary terminology.

Suppose that the projective space has been cut along a plane which is taken as the plane at infinity. Then for the set of points on every finite line (that is, a line not lying in the plane at infinity) a "betweenness" relation may be introduced (see § 90). Namely, if O_∞ is the point at infinity of a finite line a and A, B, C are three other points on this line, then C is said to lie between A and B if the pair C, O_∞ divides the pair A, B . As we know, the finite elements of space then satisfy all Hilbert's axioms of the first two groups. On the basis of these axioms the set of finite points of the line may be ordered so that if C follows A and precedes B , it lies between A and B in the sense just described. Subject to this condition, the set of finite points on a line may be ordered in just two different ways and the two orders are the reverse of each other (see § 14). Each of these orders will be called a *linear order* on the deleted projective line.*

Now we may formulate the axiom of continuity, the only one in the third group and the last of projective geometry.

Group III. Dedekind's axiom of continuity.

(III) *Let a be any deleted projective line from which the point O_∞ has been excluded. If the set of remaining points of this line is divided into two classes such that (i) every point belongs to only one class, (ii) each class contains points, and (iii) each point of the first class in one of the two linear orders on line a precedes every point of the second class, then either the first class contains a point which follows every point of this class or the second class contains a point preceding all the other points of this class. This axiom may briefly be stated as follows:*

In every Dedekind's cut of an ordered set of points of a deleted projective line only one of the two classes contains a closing element.

§ 95. The next few pages deal with the introduction of a coordinate system on the projective line.

Take three points, denoted by 0, 1 and ∞ , on a projective line a (Fig. 110). The point ∞ will be called the point at infinity and the remaining points of a will be called finite points. Assuming ∞ to be excluded from a , we introduce on this line a linear order in which point 0 precedes point 1. Next, by 2 we denote the point which together with 0 forms a pair harmonically conjugate to the pair 1, ∞ . By Theorem 9, the pair 0, 2 divides the pair 1, ∞ . There-

* By the deleted projective line we shall mean a projective line from which the point at infinity has been excluded. Similarly, the terms deleted projective plane and deleted projective space will be used to signify the projective plane and the projective space from which the line at infinity and the plane at infinity, respectively, have been excluded.

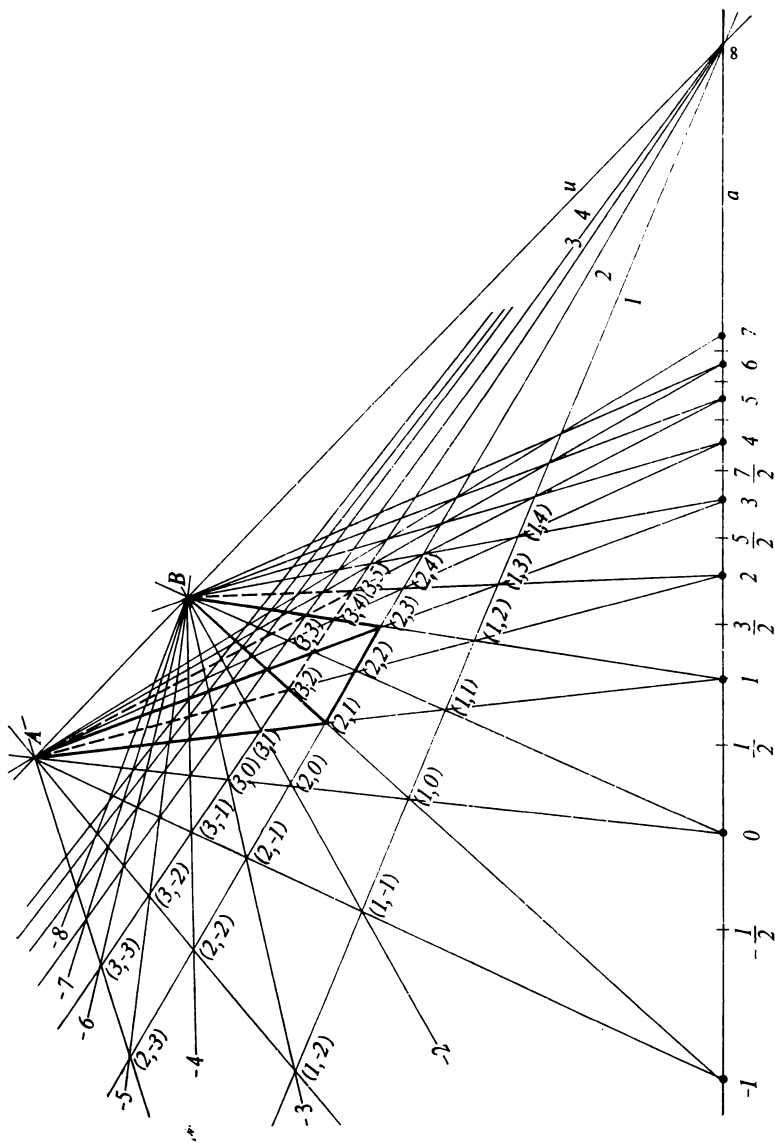


Figure 110

fore in the linear order introduced on a point 1 lies between 0 and 2; in other words, point 2 follows points 0 and 1. Next, we label by 3 the point which forms with 1 a pair harmonically conjugate to the pair 2, ∞ ; by 4 the point forming with 2 a pair harmonically conjugate to the pair 3, ∞ , and so on. Thus we obtain an infinite set of points labelled by numbers 0, 1, 2, 3, 4, $\dots$. Obviously, the point p for any $p \geq 1$ in this set follows each of the points 0, 1, 2, $\dots$, $p - 1$.

Next, we label by -1 the point which forms with 1 a pair harmonically conjugate to the pair 0, ∞ ; by -2 the point forming with 0 a pair harmonically conjugate to the pair -1 , ∞ ; and so on. In this way, we obtain the points $-m$, $-m + 1$, $\dots$, -3 , -2 , -1 , 0, 1, 2, 3, 4, $\dots$, n , $\dots$, which follow each other in the linear order defined on the deleted line a . These points will be called *the integer points of the projective scale*.

In order to construct them in the most convenient way, we proceed in the following manner.

Draw two lines, one denoted by 1 and the other by u , through point ∞ of the line a , and choose a point A on u (Fig. 110). Draw also the lines $A0$ and $A1$ joining point A with points 0 and 1. These lines intersect line 1 in two points, which we denote by $(1, 0)$ and $(1, 1)$. Next, by drawing a line through points 0 and $(1, 1)$, we find a point B where this line meets u . After this, drawing a line through points B and 1, we determine a point $(1, 2)$ on line 1 and projecting this point from centre A to line a , we obtain the point which was earlier denoted by 2, since this is the point which forms with 0 a pair harmonically conjugate to the pair 1, ∞ . To see this, it suffices to consider a complete four-point with vertices A , B , $(1, 1)$, $(1, 2)$; the points 1 and ∞ are diagonal points of this four-point and the points 0 and 2 lie on its two opposite sides, which implies the harmonic conjugateness of the pairs 0, 2 and 1, ∞ . Having found point 2, we project this point from B to line 1 to obtain the point $(1, 3)$, and projecting this last point from centre A to line a , we get the point 3; having located point 3, we project this point from B to line 1 and obtain the point $(1, 4)$, and then project this point from A to line a to get the point 4; and so on.

The integer points with the negative sign can be obtained similarly. To illustrate, if we project point $(1, 0)$ from B to line a , we obtain point -1 , and projection of this point from A to line 1 yields the point $(1, -1)$ whose projection, in turn, from B to a results in the point -2 , and so on.

By construction, two lines one of which joins B with integer point n and the other joins A with integer point $n + 1$ meet on line 1 for any n .

Besides this result, one can also show that two lines one of which joins B with an integer point n and the other joins A with integer

point $n + 2$ also intersect on a certain line for any n . This line has been denoted by number 2 in Fig. 110 and the points of intersections, on this line, of the above lines have been denoted by $\dots$, $(2, -3)$, $(2, -2)$, $(2, -1)$, $(2, 0)$, $(2, 1)$, $(2, 2)$, $\dots$.

Similarly, two lines one of which joins B with n and the other joins A with $n + 3$ intersect on a certain line 3 for any n , resulting on this line in a system of points $\dots$, $(3, -3)$, $(3, -2)$, $(3, -1)$, $(3, 0)$, $(3, 1)$, $(3, 2)$, $\dots$.

Two lines one of which joins B with n and the other joins A with $n + 4$ intersect on a certain line 4 for any n , etc.

The proof of these assertions will be given only for the system of points $\dots$, $(2, -3)$, $(2, -2)$, $(2, -1)$, $(2, 0)$, $(2, 1)$, $\dots$; this proof can be extended to other systems without any difficulty.

Thus we shall show that the points $\dots$, $(2, -3)$, $(2, -2)$, $(2, -1)$, $(2, 0)$, $(2, 1)$, $(2, 2)$, $\dots$ lie on one line.

To do this, note first that for any n the pair of points A , $(1, n)$ is harmonically conjugate to the pair of points $(2, n)$, n .

Indeed, by construction, these pairs are obtained by projecting from point B two mutually harmonic pairs of points ∞ , $n - 1$ and $n - 2$, n of a ; accordingly, by Theorem 6 of § 86, the given pairs are themselves harmonically conjugate.

Let 2 denote the line emanating from ∞ towards $(2, 0)$. As we see, two pairs of rays u , 1 and 2, a emanating from ∞ project two pairs of harmonically conjugate points A , $(1, 0)$ and $(2, 0)$, 0. Therefore the rays in question meet any line in two pairs of harmonically conjugate points (see § 86, Theorem 6).

In particular, the line joining A and n intersects the rays u , 1 and a at three points, A , $(1, n)$ and n , and the ray 2 at the point which is the fourth harmonic to the three points mentioned. But, as we know, such a point is $(2, n)$. And because the fourth harmonic to three given points is determined uniquely, the point $(2, n)$ must lie on line 2 for any n .

Having established that the points $\dots$, $(2, -3)$, $(2, -2)$, $(2, -1)$, $(2, 0)$, $(2, 1)$, $(2, 2)$, $\dots$ lie on one line, it can easily show that the points $\dots$, $(3, -3)$, $(3, -2)$, $(3, -1)$, $(3, 0)$, $(3, 1)$, $(3, 2)$, $\dots$ also lie on one line. For this, it should first be noted that for any n the pair A , $(2, n)$ is harmonically conjugate to the pair $(3, n)$, $(1, n)$. Then, using the fact that each of the two systems of points $\dots$, $(1, -2)$, $(1, -1)$, $(1, 0)$, $(1, 1)$, $(1, 2)$, $\dots$ and $\dots$, $(2, -2)$, $(2, -1)$, $(2, 0)$, $(2, 1)$, $(2, 2)$, $\dots$ lies on one line each, one may show in the above manner that the system of points $\dots$, $(3, -2)$, $(3, -1)$, $(3, 0)$, $(3, 1)$, $(3, 2)$, $\dots$ is also situated on one line. Similarly, it may be shown that the points $\dots$, $(4, -2)$, $(4, -1)$, $(4, 0)$, $(4, 1)$, $(4, 2)$, $\dots$ lie on one line, and so on. We are now in a position to prove the following auxiliary theorem which will be frequently used.

Theorem A. *If x , y and $z = (x + y)/2$ are integers, then the integer point z forms with ∞ a pair which divides the pair of integer points x , y harmonically.*

The point which forms with ∞ a pair dividing two given points P and Q harmonically on line a will be referred to as the *projective centre of the segment PQ* (a projective centre, of course, depends on ∞). Then the theorem mentioned just now may be formulated as follows:

If x , y and $z = (x + y)/2$ are integers, then the integer point z is a projective centre of the segment determined by the integer points x and y .

For the proof of this theorem, we assume that $y > x$. The given condition implies that the number $y - x$ is even. If $y - x = 2$, the conclusion of the theorem is evidently valid because with $y - x = 2$ the point $(x + y)/2$ is the projective centre of the segment xy since this fact is the basis of the definition of the projective scale. The construction shown in Fig. 110 is based on this very fact; one may find from Fig. 110 that the lines joining A with y and B with x intersect line 1 at two points which together with A and B are vertices of a four-point with diagonal points $(x + y)/2$ and ∞ and the pair of opposite sides passing through the points x and y . And this means that $(x + y)/2$ is the projective centre for the segment xy . The theorem may be verified in the same way for $y - x = 4$, $y - x = 6$, etc.

For instance, suppose $y - x = 4$. Consider the lines joining A with y and B with x and the points of intersection of these lines with line 2. These points of intersection together with A and B are vertices of a four-point with diagonal points $(x + y)/2$ and ∞ and the pair of opposite sides passing through x and y . And this means that the point $(x + y)/2$ is the projective centre for the segment xy . In Fig. 110, a four-point is represented in heavy lines whose examination reveals that the point $1 = (3 + (-1))/2$ is the projective centre of the segment with endpoints 3 and -1 ; in this case, $y - x = 3 - (-1) = 4$.

When $y - x = 6$, the theorem is verified in the same way, only now one has to use line 3, and when $y - x = 8$, line 4 is to be used, and so on.

In Fig. 110, a heavy dashed line shows a four-point. Upon examination one finds that the point $2 = (5 + (-1))/2$ is the projective centre of the segment with endpoints 5 and -1 , and in this case $y - x = 5 - (-1) = 6$.

So far we have considered only integer points. Now we shall "condense" the projective scale by introducing new points labeled by fractions. $\frac{1}{2}$

First we find the projective centre of the segment $(0, 1)$ and call this centre $\frac{1}{2}$. Then, starting with three points $0, \frac{1}{2}, \infty$, we construct a projective scale in the same way as we did by starting with the points $0, 1, \infty$. We thus obtain a set of points fulfilling the role of integer points in the new scale; these points are labeled $\dots, -\frac{4}{2}, -\frac{3}{2}, -\frac{2}{2}, -\frac{1}{2}, 0, \frac{1}{2}, \frac{2}{2}, \frac{3}{2}, \frac{4}{2}, \dots$. It is easy to see that all the points of the form $\dots, -\frac{4}{2}, -\frac{2}{2}, 0, \frac{2}{2}, \frac{4}{2}, \dots$ in the new scale coincide respectively with the points $\dots, -2, -1, 0, 1, 2, \dots$ in the old scale. Indeed, the point $\frac{2}{2}$ of the new scale coincides with the point 1 of the old scale, since the segments $(0, \frac{2}{2})$ and $(0, 1)$ have a common projective centre, namely the point $\frac{1}{2}$; the point $\frac{4}{2}$ of the new scale coincides with point 2 of the old as the segments $(0, \frac{4}{2})$ and $(0, 2)$ have a common projective centre. In fact, Theorem A implies that the segment $(0, \frac{4}{2})$ has for projective centre the point $\frac{2}{2}$ of the new scale, and the same theorem shows that the projective centre of the segment $(0, 2)$ is the point 1 of the old scale, but, as just noted, the points $\frac{2}{2}$ and 1 coincide. Further, the point $\frac{6}{2}$ of the new scale coincides with the point 3 of the old because the segments $(\frac{2}{2}, \frac{6}{2})$ and $(1, 3)$ have a common vertex (we already know that the points $\frac{2}{2}$ and 1 coincide) and a common projective centre which is point $\frac{4}{2}$ of the new scale and simultaneously point 2 of the old scale. Continuing in this manner, we establish that the point of the form $\frac{2n}{2}$ and n coincide and similarly that the points $-\frac{2n}{2}$ and $-n$ coincide.

Now it is clear that all the points of the new scale, that is, of the scale determined by the points $0, \frac{1}{2}, \infty$, may be obtained by adjoining to the points of the old scale the projective centres of the segments of the form $(n, n+1)$.

Moreover, the following generalization of Theorem A may be formulated.

Theorem B. *Given any two integers x and y , the number $z = (x + y)/2$ always determines the projective centre of the segment xy .*

The condensation of the scale which was carried out above by adjoining to integer points the projective centres of the segments $(n, n + 1)$ may be continued beyond the first step. Considering the points of the form $\frac{n}{2}$ (which includes all integer points in the form of fractions with common factors), we obtain the projective centre of each segment $(\frac{n}{2}, \frac{n+1}{2})$ and label it by the number $(\frac{n}{2} + \frac{n+1}{2})/2 = \frac{2n+1}{2^2}$. Thus we shall obtain points determined by numbers of the form $\frac{n}{2^2}$, which are already present in the scale; applying the same technique to the new points, we obtain points determined by the numbers of the form $\frac{n}{2^3}$, and so on.

Thus, whatever the binary fraction $\frac{n}{2^m}$, there is one completely determined point on the deleted projective line a which is labeled $\frac{n}{2^m}$ in our construction. On the basis of the preceding result, we may assert that for any given m the points of the form $\frac{n}{2^m}$ ($n = \dots, -2, -1, 0, 1, 2, \dots$) constitute the projective scale determined by the points $0, \frac{1}{2^m}$ and ∞ . Thus we have the following generalization of Theorem B.

Theorem C. *Given any two binary fractions x and y , the number $z = (x + y)/2$ is always the projective centre of the segment xy .*

In fact, putting $x = \frac{n}{2^p}$ and $y = \frac{m}{2^q}$ we first reduce them to fractions with a common denominator and then represent them in the form $x = \frac{M}{2^r}$ and $y = \frac{N}{2^r}$. Now x and y can be treated as integer points of the scale determined by the points $0, \frac{1}{2^r}$ and ∞ . Then Theorem C is evidently a direct corollary of Theorem B.

§ 96. We shall now show that the points labeled by binary fractions (henceforth they will be referred to as rational-binary) everywhere cover the projective line a densely.

The proof will use the "method of contradiction". Suppose that a segment PQ does not contain rational-binary points and suppose, for the sake of definiteness, that P precedes Q in the linear order introduced on the deleted projective line a . From the assumptions made, there arise three cases:

(1) there are rational-binary points which precede P and rational-binary points which follow Q ;

(2) there are rational-binary points preceding P but there are no rational-binary points following Q ;

(3) there are rational-binary points following Q but no points preceding P .

Assuming that the segment PQ is free of rational-binary points, we have to show that all the three cases lead to contradictions.

Consider the first case.

We divide all the points of the deleted projective line a into two classes, assigning to the second class each rational-binary point which follows Q and every point of a which follows this point and to the first class all the remaining points. Evidently, such a classification is a Dedekind's cut. According to Axiom III, there is a point performing this cut, that is, a point which closes one of the classes; call this point Q_0 . First, it is easy to see that Q_0 cannot precede Q . What is more, if Q and Q_0 are distinct, then between these points there is no rational-binary point, since otherwise Q_0 would belong to the second class but would not be its first point (would not close this class). At the same time, on a every neighbourhood of Q_0 contains rational-binary points. In fact, if there is a neighbourhood of Q_0 that does not contain rational-binary points, then all the points of this neighbourhood, including Q_0 , would belong to the first class and the point Q_0 would not be the last point of this class (would not close the class). Observe also that Q_0 cannot coincide with ∞ because, according to the hypothesis, there are rational-binary points which follow Q and necessarily separate Q_0 from ∞ .

Now we divide all the points of the deleted projective line a into two classes assigning this time to the first class each rational-binary point preceding P as well as any other rational-binary point of a which precedes this rational-binary point, and assign to the second class all the remaining points of a . This again results in a Dedekind's cut; call the point performing this cut P_0 . As in the above discussion, it can be established, first, that P_0 cannot follow P and that there are no rational-binary points between P_0 and P provided these points are distinct, secondly, that every neighbourhood of P_0 contains rational-binary points and, finally, that P_0 cannot coincide with ∞ .

Thus the segment P_0Q_0 , like the segment PQ , does not contain rational-binary points but such points are contained in any neighbourhoods of P_0 and Q_0 .

Let X, Y be any two points, distinct from ∞ , of line a and $Z = f(X, Y, \infty)$ be the point which together with ∞ forms a pair Z, ∞ that divides the pair X, Y harmonically. The point Z is simply the projective centre of the segment XY . Suppose further that $R_0 = f(P_0, Q_0, \infty)$ is the projective centre of the segment P_0Q_0 . We know that the segment P_0Q_0 contains the point R_0 . By Theorem

10, the function $f(X, Y, \infty)$ with $X = P_0, Y = Q_0$ is continuous. Therefore there are neighbourhoods Δ_1 and Δ_2 of points P_0 and Q_0 such that, once X lies inside Δ_1 and Y inside Δ_2 , the point $Z = f(X, Y, \infty)$ lies inside the segment P_0Q_0 . According to the preceding result, Δ_1 and Δ_2 contain rational-binary points. If x is the binary fraction corresponding to a certain X inside Δ_1 and y is the binary fraction corresponding to Y inside Δ_2 , then $Z = f(X, Y, \infty)$ is, according to Theorem C, the rational-binary point corresponding to the binary fraction $\frac{x+y}{2}$. Therefore the

segment P_0Q_0 necessarily contains a rational-binary point. But, by its construction, this segment is free of any rational-binary point. Thus, assuming that there is a segment PQ that does not contain rational-binary points, we arrive at a contradiction for the first of the three cases mentioned.

We now turn to the second case.

We have to show, essentially, that rational-binary points cannot all precede a given point P of the deleted projective line. Assuming the contrary, we divide all the points of the deleted line into two classes. To the first class we assign each rational-binary point as well as any other point which precedes a rational-binary point, and all the remaining points are assigned to the second class. Thus we obtain a Dedekind's cut; by Axiom III, there is a point P_0 performing this cut. As in the preceding case, it can first be shown that if P_0 and P are distinct, then there are no rational-binary points between them; in other words, there are no rational-binary points on the entire line from P_0 to ∞ , and, secondly, that every neighbourhood of P_0 contains rational-binary points.

This immediately leads to a contradiction.

Indeed, let X be any point of the deleted line and Y a point which is determined by X so that the pair O, Y divides the pair X, ∞ harmonically. Employing the notation introduced above, we write $Y = f(X, \infty, 0)$. Set $R_0 = f(P_0, \infty, 0)$; this point lies inside the segment (P_0, ∞) because 0 precedes P_0 . By Theorem 10, the function $Y = f(X, \infty, 0)$ is continuous for $X = P_0$. Therefore there is a neighbourhood Δ of P_0 such that if X lies inside Δ , the point Y will lie inside the segment (P_0, ∞) . The neighbourhood Δ contains, like any other neighbourhood of P_0 , rational-binary points. Let x be the binary fraction corresponding to a point X inside Δ and let $y = 2x$ be the binary fraction corresponding to the point Y . According to Theorem C, point X is the projective centre of the segment $(0, Y)$ and therefore Y is related to X by $Y = f(X, \infty, 0)$. But since X lies inside Δ , point Y lies inside the segment (P_0, ∞) . Thus this segment contains a rational-binary point, contrary to its construction. Therefore the second of the three possible cases mentioned above must be rejected.

There is no need for treating the third case separately because in essence this case is not different from the preceding case. This completes the proof of our assertion.

§ 97. We see that rational-binary points everywhere cover the projective line densely. But they do not exhaust all the points of the projective line. There are infinitely many other points, with which we shall associate, according to some rule, real numbers other than binary fractions.

Let M be any point of the deleted projective line. Let $\{P\}$ and $\{Q\}$ be the sets of rational-binary points preceding and following M , respectively; if M itself is a rational-binary point, then it will be included, for example, in $\{P\}$. Further, let $\{p\}$ and $\{q\}$ denote the sets of binary fractions corresponding, respectively, to the points of $\{P\}$ and $\{Q\}$. From the preceding results it follows that (a) if p and q are any fractions of $\{p\}$ and $\{q\}$, respectively, then $p < q$; and (b) the sets $\{p\}$ and $\{q\}$ together constitute the entire set of rational-binary fractions.

Therefore there is a unique number x which is greater than every member of $\{p\}$ * and smaller than every member of $\{q\}$. This is the number which we associate with M .

Thus every point of the deleted projective line has corresponding to it a completely defined number which will be referred to as the *projective coordinate* of the point in question.

The correspondence associating a definite coordinate with every point (except ∞) has the following properties.

(1) To distinct points there correspond distinct coordinates; moreover, if a point M_1 with coordinate x_1 precedes another point M_2 with coordinate x_2 , then $x_1 < x_2$.

Indeed, since the set of rational-binary points on the projective line is everywhere dense, there must exist a rational-binary point P with coordinate p between M_1 and M_2 such that $x_1 < p < x_2$.

(2) Given a real number x , there is a point with coordinate x .

Indeed, if x is a binary fraction, then, as known from an earlier result, there is a rational-binary point which has for its coordinate this number x . But if x is any other real number, then, to prove the assertion, we divide all the binary fractions into two sets, $\{p\}$ and $\{q\}$. To set $\{p\}$ we assign each binary fraction p if $p < x$, and to set $\{q\}$ we assign each binary fraction q if $x < q$. Simultaneously, we may treat the set of rational-binary points as divided into two sets, $\{P\}$ and $\{Q\}$, which are composed of the points with coordinates, respectively, from $\{p\}$ and $\{q\}$. Next, we perform a Dedekind's cut over the set of all points of the deleted projective line by assigning to the first class of the cut every point of $\{P\}$ and every other

* Or x is the largest of these numbers if M is a rational-binary point.

point of the line if this precedes a point of $\{P\}$; to the second class we assign all the remaining points.

By Axiom III, there is a point M performing this cut. Evidently, M follows every point of $\{P\}$ and precedes every point of $\{Q\}$. Therefore the coordinate of M must be greater than every fraction of $\{p\}$ and less than every fraction of $\{q\}$, but such a number is only the given number x . Hence the coordinate of M is x .

(3) The correspondence between the points of the deleted projective line and their coordinates is continuous, that is, if the sequence of points M_n has the limit point M , then the coordinate x of M will be the limit of the sequence of coordinates x_n of points M_n , and vice versa. In short, $x_n \rightarrow x$ follows from $M_n \rightarrow M$, and $M_n \rightarrow M$ follows from $x_n \rightarrow x$. This property is the consequence of the fact that (i) the correspondence between the points of the deleted number axis and their coordinates is one-to-one, (ii) every real number is the coordinate of some point, and (iii) the order of points is the same as the order of their coordinates. In view of these facts, if x is the coordinate of M , then to every neighbourhood of M on the deleted projective line there corresponds a neighbourhood of coordinate x on the number axis, and, conversely, to every neighbourhood of coordinate x on the number axis there corresponds a neighbourhood of M on the deleted projective line. Thus, if M_n falls inside a neighbourhood of M , then x_n falls inside the corresponding neighbourhood of x ; conversely, if x_n falls inside a neighbourhood of x , then M_n falls inside the corresponding neighbourhood of M . This means that if $M_n \rightarrow M$, then $x_n \rightarrow x$, and if $x_n \rightarrow x$, then $M_n \rightarrow M$.

(4) If M_1 and M_2 are any two points with coordinates x_1 and x_2 , then the projective centre of the segment M_1M_2 has the coordinate $\frac{x_1+x_2}{2}$.

To prove this statement, we take two sequences of binary fractions $p_1^{(n)}$ and $p_2^{(n)}$ converging, respectively, to x_1 and x_2 , so that $\lim_{n \rightarrow \infty} p_1^{(n)} = x_1$ and $\lim_{n \rightarrow \infty} p_2^{(n)} = x_2$. Denote by $P_1^{(n)}$ and $P_2^{(n)}$ the rational-binary points with coordinates $p_1^{(n)}$ and $p_2^{(n)}$, by $c^{(n)}$ the coordinate of the projective centre of the segment $P_1^{(n)}P_2^{(n)}$ and by c the coordinate of the projective centre of the segment M_1M_2 . Theorem 10 implies that $c = \lim_{n \rightarrow \infty} c^{(n)}$. On the other hand, by Theorem C (which asserts the same fact regarding rational-binary points what we are going to establish now for arbitrary points), we have $c^{(n)} = \frac{p_1^{(n)}+p_2^{(n)}}{2}$. Hence $c = \lim_{n \rightarrow \infty} \frac{p_1^{(n)}+p_2^{(n)}}{2} = \frac{x_1+x_2}{2}$, as required.

The first three properties of the projective coordinate system, just proved, may be combined thus: the projective coordinates are constructed by means of a one-to-one and continuous correspondence between the set of all points of the deleted projective line and the

set of all real numbers; moreover, this correspondence is such that the points of the line and their corresponding numbers (the coordinates) are in the same order. It should be noted that these properties are possessed, besides the aforementioned (projective) system of coordinates, by many other coordinate systems.

In contrast, the fourth property is characteristic of this system and was taken as the basis of its definition right from the beginning. In other words, among all the possible coordinate systems on a deleted projective line the projective system is distinguished by the fact that in this system the coordinate of the projective centre is always the arithmetic mean of the coordinates of the endpoints.

In concluding this section, we emphasize that the projective system is defined by means of three points: 0 , 1 , ∞ . By changing these points, we can obtain various other projective systems of coordinates on the same line.

5.6. Projective coordinate systems on a plane and in space

§ 98. Let a line be given in any projective plane. Denoting this line by ∞ , we shall call it the line at infinity and assume the projective plane to be cut along this line.

We shall now present a technique for associating in a one-to-one manner a pair of real numbers with points of the deleted projective plane. These numbers will be called projective coordinates of the corresponding points.

The projective coordinate system is determined by prescribing the following geometrical elements: a point O , called the origin of the coordinate system, two lines through O , one of which is called the x -axis and the other the y -axis and, finally, a point E (not lying on these axes).

Let ∞_x and ∞_y be the points at infinity of the x - and y -axes, that is, their points of intersection with line ∞ (Fig. 111). Project point E from ∞_y to the x -axis and from ∞_x to the y -axis. Each of the projections thus obtained is labeled by number 1 . Then, as in the preceding section, on the x -axis we introduce a linear system of projective coordinates determined by three points 0 , 1 , ∞_x ; similarly, a coordinate system determined by points 0 , 1 , ∞_y is introduced on the y -axis.

Consider an arbitrary point M of the deleted projective plane. Let M_x denote the projection of M from ∞_y to the x -axis and M_y the projection of M from ∞_x to the y -axis. The point M_x has a certain coordinate x in the linear coordinate system introduced on the x -axis; similarly, M_y has a coordinate y on the y -axis. The numbers x and y will be called the *projective coordinates of point M on the plane*.

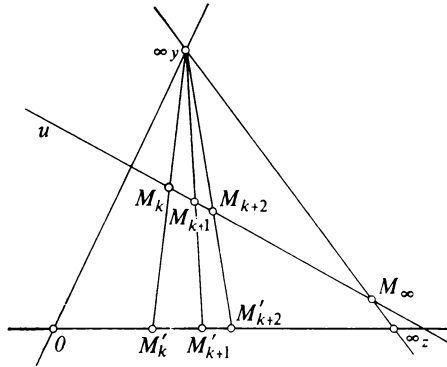


Figure 112

Similarly,

$$\frac{y_k + y_{k+2}}{2} = y_{k+1} \quad (**)$$

Now let M be any point of the plane having coordinates x, y and let $L(M) = Ax + By$ be a linear function of this point. Choose numbers A, B, C so that

$$\left. \begin{aligned} L(M_0) + C &= Ax_0 + By_0 + C = 0 \\ L(M_1) + C &= Ax_1 + By_1 + C = 0 \end{aligned} \right\} \quad (***)$$

The relations (*) and (**) yield

$$L(M_k) + L(M_{k+2}) - 2L(M_{k+1}) = 0$$

Hence, if $L(M_k) + C = 0$ and $L(M_{k+1}) + C = 0$, then $L(M_{k+2}) + C = 0$. This together with (***) yields

$$L(M_n) + C = Ax_n + By_n + C = 0$$

for any integer point M_n . Thus the coordinates of all integer points of line u satisfy an equation of the first degree

$$Ax + By + C = 0$$

It is not difficult to show that this equation is satisfied not only by the coordinates of integer points but also by the coordinates of all rational-binary points. In fact, the rational-binary points were introduced by successively "condensing" the projective scale; the set of all integer points together with projective centres of the segments determined by these points will be called a "first condensation" of the scale, and the set of all points of a "first condensation" together

with the projective centres of segments determined by these points as a "second condensation" of the scale, and so on.

If the point $M_{k+1/2}$ is the projection centre of the segment $M_k M_{k+1}$, its coordinates satisfy the relations

$$x_{k+1/2} = \frac{x_k + x_{k+1}}{2}, \quad y_{k+1/2} = \frac{y_k + y_{k+1}}{2}$$

which yields

$$L(M_k) + L(M_{k+1}) - 2L(M_{k+1/2}) = 0$$

Therefore, if $L(M_k) + C = 0$ and $L(M_{k+1}) + C = 0$, then we must have $L(M_{k+1/2}) + C = 0$. Thus the equation $Ax + By + C = 0$ is satisfied by the coordinates of all points of the first condensation. Similar arguments show that the same equation is satisfied by the coordinates of points of the second condensation, and so on. Since the function $Ax + By + C = 0$ is continuous and is zero at points of the set which is everywhere dense on u , this function vanishes generally at all points of line u . In other words, the coordinates of every point of u satisfy the equation $Ax + By + C = 0$; at the same time, it is clear that the coordinates of points on u exhaust all the pairs (x, y) of solutions of this equation. Accordingly, this equation is the equation representing line u , which, as we find, is an algebraic equation of the first degree. ■

Besides the equation

$$Ax + By + C = 0$$

which will be referred to as the *general equation of a line*, we shall also use the following special forms of this equation:

(1) The equation solved for one of the coordinates (say, y)

$$y = kx + l$$

In this form the equation is identical to the slope-intercept form, which is frequently used in analytic geometry on the Euclidean plane. But, of course, now k cannot be called the slope of the line since all the metric concepts, including the magnitude of an angle, are absent in projective geometry.

(2) The equation involving the coordinates x_0, y_0 of one of the points on the line (point-slope form):

$$y - y_0 = k(x - x_0)$$

(3) The equation

$$\frac{x}{a} + \frac{y}{b} = 1$$

In appearance, this equation is identical to the intercept form, which is well known in elementary analytic geometry. But we must not interpret a, b as the lengths of the segments intercepted by the line

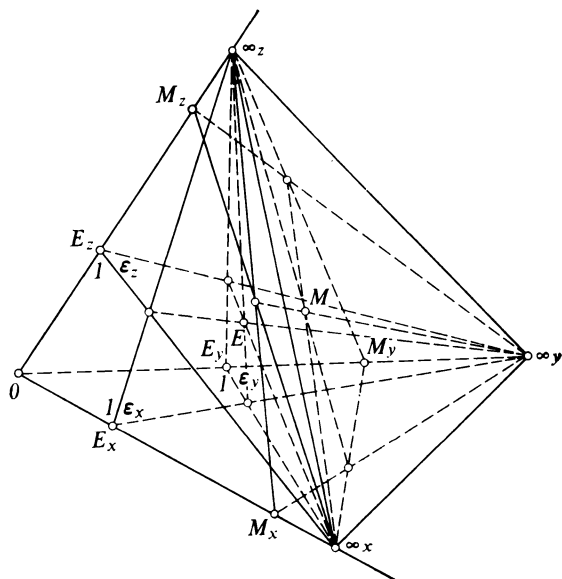


Figure 113

on the axes but rather as the projective coordinates of the points of intersection of the line with the axes.

§ 99. Now the projective coordinates in space will be dealt with.

Let a plane be given in the projective space. Denoting this plane by ∞ , we shall call it the plane at infinity and assume that the space has been cut along this plane.

In the deleted projective space, we introduce a coordinate system by prescribing a point O , to be called the origin, three lines through O , to be called, respectively, the x -, y - and z -axis, and a point E not lying in any of the three planes determined by the axes taken two at a time (Fig. 113).

Let ∞_x , ∞_y , ∞_z be the points at infinity of the axes, that is, the points of intersection of these axes with the plane ∞ . Denote by ϵ_x , ϵ_y , ϵ_z the planes determined, respectively, by the points E , ∞_y , ∞_z ; E , ∞_x , ∞_z ; and E , ∞_x , ∞_y . The planes ϵ_x , ϵ_y , ϵ_z intersect the x -, y - and z -axis at some points E_x , E_y , E_z , respectively; we label each of these points by I .

Then we introduce a linear system of projective coordinates determined by three points O , I , ∞_x on the x -axis, and, similarly, introduce coordinate systems on the y - and z -axis, respectively, by means of the points O , I , ∞_y and O , I , ∞_z .

Consider an arbitrary point M of the deleted projective space. Let M_x , M_y , M_z denote, respectively, points of intersection of the

plane $M\infty_y\infty_z$ and the x -axis, the plane $M\infty_x\infty_z$ and the y -axis, and the plane $M\infty_x\infty_y$ and the z -axis. Then in the linear coordinate system on the x -axis the point M_x has a certain coordinate x , and, similarly, M_y on the y -axis and M_z on the z -axis have, respectively, the coordinates y and z .

The numbers x, y, z will be called the *projective coordinates of point M in the space*. We shall now establish a fundamental property of projective coordinates in the form of the following

Theorem 13. *In projective coordinates, every plane is represented by a first degree algebraic equation.*

Proof. In order to simplify the proof of this theorem, we shall examine only those planes which do not pass through the origin or the points $\infty_x, \infty_y, \infty_z$.

Consider a plane α which meets the coordinate axes at points A, B, C , respectively. If the points A, B, C have coordinates a, b, c , respectively, then, by the condition imposed, we have $a \neq 0, b \neq 0, c \neq 0$. We shall show that α is represented by the equation

$$\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1 \quad (*)$$

Take any point M in α and let its coordinates be $\bar{x}, \bar{y}, \bar{z}$. Denote by μ the plane determined by the points M, ∞_y, ∞_z , by R the point of intersection of μ and the x -axis, by P and Q the points where the line of intersection of the planes α and μ meets the planes Oxy and Oxz (Fig. 114). Obviously, the line PQ contains point M .

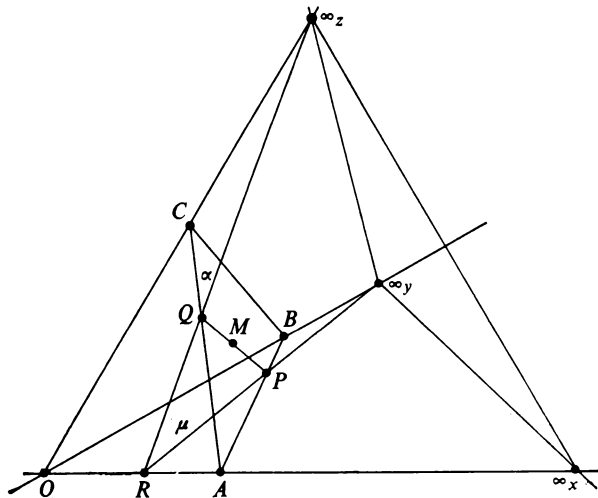


Figure 114

In the projective coordinate system on the plane $R\infty_y\infty_z$ the line PQ is represented by an equation of the form

$$\frac{y}{p} + \frac{z}{q} = 1$$

This equation must be satisfied by the coordinates $y = \bar{y}$, $z = \bar{z}$ because M lies on PQ ; accordingly,

$$\frac{\bar{y}}{p} + \frac{\bar{z}}{q} = 1 \quad (**)$$

We now determine the parameters p and q . The equation of the line AB in projective coordinates on the plane Oxy is

$$\frac{x}{a} + \frac{y}{b} = 1$$

But on the plane Oxy the projective coordinates of P are the numbers $\bar{x}$, p (in the space, the coordinates of P are $\bar{x}$, p , O). Therefore

$$\frac{\bar{x}}{a} + \frac{p}{b} = 1$$

which yields

$$p = b \left(1 - \frac{\bar{x}}{a} \right)$$

Similarly, the equation

$$\frac{x}{a} + \frac{z}{c} = 1$$

representing the line AC on the plane Oxz , gives, with $x = \bar{x}$, $z = q$,

$$q = c \left(1 - \frac{\bar{x}}{a} \right)$$

For these values of p and q equation $(**)$ yields

$$\frac{\bar{y}}{b(1-\bar{x}/a)} + \frac{\bar{z}}{c(1-\bar{x}/a)} = 1$$

or

$$\frac{\bar{x}}{a} + \frac{\bar{y}}{b} + \frac{\bar{z}}{c} = 1$$

Thus the coordinates of any point of plane α satisfy equation $(*)$.* ■

The derivation of the special forms of the equation of the plane when it contains the origin or the points at infinity of the axes is

* We should have also shown that the coordinates of a point not on α do not satisfy equation $(*)$, but this result immediately follows from the fact that equation $(*)$ can be solved uniquely for each of the coordinates.

left to the reader. All these cases are included in the general form

$$Ax + By + Cz + D = 0$$

Since a plane is determined by a first degree equation, a line in space can be determined by a pair of equations

$$\begin{cases} A_1x + B_1y + C_1z + D_1 = 0 \\ A_2x + B_2y + C_2z + D_2 = 0 \end{cases}$$

By means of some algebra these equations may be reduced to the "canonical" form

$$\frac{x-x_0}{m} = \frac{y-y_0}{n} = \frac{z-z_0}{p}$$

where x_0, y_0, z_0 are the coordinates of some point on the line.

§ 100. Hitherto we have examined a coordinate system on the deleted projective line, on the deleted projective plane and in the deleted projective space. In other words, when considering the projective line, we associated coordinates with every point of this line, but one point (denoted by ∞) was assigned no coordinate. In the projective plane and projective space with every point a pair and a triple of coordinates were respectively associated, but the points of a certain line and, in the case of space, points of a certain plane (both denoted by ∞) were not assigned any coordinates.

In order to arithmetize the entire projective line, projective plane and projective space, one has to employ the homogeneous coordinates. First, we shall discuss homogeneous coordinates on a projective line.

Let a projective line a be given. Take any three points on this line, labeling two of them by the numbers 0 and 1 and the third by the symbol ∞ . Then on line a introduce a projective coordinate system by means of the points 0, 1, ∞ . In this system every point of the line, except ∞ , has a certain coordinate. Let M be any point of a having coordinate x ; two numbers x_1 and x_2 , which are not both zero, will be called *homogeneous coordinates* of M if the ratio $x_1 \div x_2$ equals x . To point ∞ we assign homogeneous coordinates x_1 and x_2 with $x_2 = 0$. The system of homogeneous coordinates thus constructed has the following properties:

(1) every point of the projective line has homogeneous coordinates;
 (2) if x_1 and x_2 are homogeneous coordinates of M , then ρx_1 and ρx_2 , where ρ is any number $\neq 0$, are also homogeneous coordinates of M ;

(3) to distinct points there always correspond unequal ratios of their homogeneous coordinates;

(4) if $x_1 \rightarrow x_1^0$ and $x_2 \rightarrow x_2^0$, then the variable point M with homogeneous coordinates x_1, x_2 tends to the point M^0 with homogeneous coordinates x_1^0, x_2^0 .

In working with homogeneous coordinates it is especially important to appreciate the meaning of the second property. Namely, every point of the projective line has infinitely many pairs of homogeneous coordinates, which are, consequently, not determined by the corresponding point; only the ratios of the coordinates are determined. By a proper choice of ρ one of the coordinates $\rho x_1, \rho x_2$ may be made equal to any given number (for example, equal to unity). For instance, homogeneous coordinates of points 0 and ∞ , which will now be denoted by A_1 and A_2 , are the pairs of numbers (0, 1) and (1, 0). For homogeneous coordinates of point 1, henceforth denoted by E , one may take the pair (1, 1). Obviously, the homogeneous coordinates on the projective line are determined by the points A_1 (0, 1), A_2 (1, 0) and E (1, 1).

§ 101. In order to arithmetize the entire projective plane, we first introduce a system of nonhomogeneous projective coordinates on the plane by prescribing an origin O , axes Ox and Oy , a unit point E not on the axes and the line at infinity ∞ (the points at infinity of the axes are denoted by ∞_x and ∞_y). Then projective coordinates are assigned to every point, except those on ∞ , of the plane.

Now we introduce homogeneous coordinates on the projective plane; first this will be done for points which do not lie on line ∞ . The homogeneous coordinates of a point M not lying on line ∞ are three numbers x_1, x_2, x_3 , which are not all zero, such that $x_1 \div x_3 = x$ and $x_2 \div x_3 = y$, where x and y are projective (non-homogeneous) coordinates of M . When M lies on line ∞ , it has no nonhomogeneous coordinates and the definition of homogeneous coordinates just given is inapplicable for such a point. The homogeneous coordinates of a point M_∞ lying on line ∞ are three numbers x_1, x_2, x_3 subject to the following conditions:

- (i) $x_3 = 0$;
- (ii) at least one of the numbers x_1, x_2 is not zero;
- (iii) the ratio $x_1 \div x_2$ equals the ratio $B \div -A$, where A and B are the coefficients in the equation

$$Ax + By + C = 0$$

of any line passing through M_∞ , that is, x_1 and x_2 must satisfy

$$Ax_1 + Bx_2 = 0$$

Let us show the validity of the third condition, that is, that the ratio $B \div -A$ does not depend on the choice of the line passing through M_∞ . Let

$$\left. \begin{aligned} A_1x + B_1y + C_1 &= 0 \\ A_2x + B_2y + C_2 &= 0 \end{aligned} \right\} \quad (*)$$

be the equations of two lines passing through M_∞ . Since these two lines meet only at M_∞ and this point, being on line ∞ , has no num-

bers as its nonhomogeneous coordinates, the two equations (*) must be inconsistent. Therefore, necessarily

$$\begin{vmatrix} A_1 & B_1 \\ A_2 & B_2 \end{vmatrix} = 0$$

whence $B_1 \div -A_1 = B_2 \div -A_2$, showing that the third condition is indeed valid.

From the definition of homogeneous coordinates it follows that whatever the point M lying on the line $Ax + By + C = 0$, its homogeneous coordinates satisfy the relation $Ax_1 + Bx_2 + Cx_3 = 0$. This relation will be referred to as the *equation of a line in homogeneous coordinates*; replacing the coefficients A, B, C by u_1, u_2, u_3 , the equation can be written in the form

$$u_1x_1 + u_2x_2 + u_3x_3 = 0$$

(this equation does not contain any free term, a feature characteristic of homogeneous coordinates).

The basic properties of homogeneous coordinates on the projective plane are analogous to those of homogeneous coordinates on the projective line. Namely,

(1) Every point of the projective plane has homogeneous coordinates.

(2) If x_1, x_2, x_3 are homogeneous coordinates of a point M , then $\rho x_1, \rho x_2, \rho x_3$ (ρ is any number $\neq 0$) are also homogeneous coordinates of M .

(3) To distinct points there always correspond unequal ratios $x_1 \div x_2 \div x_3$ of their homogeneous coordinates.

(4) If $x_1 \rightarrow x_1^0, x_2 \rightarrow x_2^0, x_3 \rightarrow x_3^0$, then the variable point M with coordinates x_1, x_2, x_3 tends to the point M^0 with coordinates x_1^0, x_2^0, x_3^0 .

It is important to emphasize that for no point are all the three homogeneous coordinates simultaneously zero. Any one of the three nonzero coordinates $\rho x_1, \rho x_2, \rho x_3$ may be made equal to unity by a proper choice of ρ . For example, for homogeneous coordinates of point O one may take the three numbers 0, 0, 1; of point ∞_x the numbers 1, 0, 0; of point ∞_y the numbers 0, 1, 0; and of point E the numbers 1, 1, 1. Henceforth instead of ∞_x, ∞_y and 0 we shall write A_1, A_2, A_3 and call them the *vertices of the coordinate trihedron*. Evidently, the coordinate trihedron $A_1(1, 0, 0), A_2(0, 1, 0), A_3(0, 0, 1)$ and the unit point $E(1, 1, 1)$ determine a system of homogeneous coordinates on the projective plane. The choice of these four points is subject to the sole condition that no three of them are collinear.

The line A_1A_2 (earlier denoted by ∞) of the projective plane contains points whose third coordinate is zero. The equation $x_3 = 0$

is nothing but the equation of the line A_1A_2 . Similarly, the equations of line A_2A_3 and A_1A_3 are $x_1 = 0$ and $x_2 = 0$, respectively.

§ 102. The introduction of homogeneous coordinates in the projective space is accomplished along the same lines. First, a system of nonhomogeneous projective coordinates is defined with the aid of axes Ox, Oy, Oz and plane ∞ . In this system, all points of the space except those of plane ∞ have coordinates assigned to them. Then homogeneous coordinates may be defined. If a point M does not lie in plane ∞ , then its homogeneous coordinates are any four numbers x_1, x_2, x_3, x_4 , not all simultaneously zero, such that $x_1 \div x_4 = x, x_2 \div x_4 = y, x_3 \div x_4 = z$, where x, y, z are nonhomogeneous coordinates of M .

If a point M_∞ lies in plane ∞ , then its homogeneous coordinates x_1, x_2, x_3, x_4 are determined by the following conditions:

- (i) $x_4 = 0$,
- (ii) at least one of the numbers x_1, x_2, x_3 is not zero;
- (iii) the ratio* $x_1 \div x_2 \div x_3$ equals the ratio $m \div n \div p$, where m, n, p are parameters in the equations

$$\frac{x-x_0}{m} = \frac{y-y_0}{n} = \frac{z-z_0}{p}$$

of any line passing through M_∞ .

Let us show the validity of the third condition, that is, we shall show that the ratio $m \div n \div p$ does not depend on the line passing through M_∞ .

Let

$$\frac{x-x_0}{m} = \frac{y-y_0}{n} = \frac{z-z_0}{p} \quad (*)$$

and

$$\frac{x'-x'_0}{m'} = \frac{y'-y'_0}{n'} = \frac{z'-z'_0}{p'} \quad (**)$$

be the equations of two lines passing through point M_∞ of plane ∞ . Since both lines pass through one point, there must exist a plane

$$Ax + By + Cz + D = 0 \quad (\alpha)$$

containing both lines. This fact imposes some condition of an analytic nature on the parameters of equations (*) and (**). In order to obtain this condition, denote each of the equal ratios (*) by t and each of the equal ratios (**) by t' . Then equations (*) and (**) may be written in the form of the following two systems of parametric equations

$$\begin{aligned} x &= x_0 + mt & x' &= x'_0 + m't' \\ y &= y_0 + nt & y' &= y'_0 + n't' \\ z &= z_0 + pt & z' &= z'_0 + p't' \end{aligned} \quad (\beta)$$

* See the footnote at the beginning of § 71.

If the first line lies in plane (α) , the coordinates of every point on this line must satisfy the equation of this plane. Therefore the relation

$$Ax + By + Cz + D = (Am + Bn + Cp)t + (Ax_0 + By_0 + Cz_0 + D) = 0$$

must hold for any t . Accordingly,

$$Am + Bn + Cp = 0, \quad Ax_0 + By_0 + Cz_0 + D = 0$$

Similarly, since the second line also lies in plane (α) ,

$$Am' + Bn' + Cp' = 0, \quad Ax'_0 + By'_0 + Cz'_0 + D = 0$$

Thus we obtain a system of equations

$$\left. \begin{aligned} A(x_0 - x'_0) + B(y_0 - y'_0) + C(z_0 - z'_0) &= 0 \\ Am + Bn + Cp &= 0 \\ Am' + Bn' + Cp' &= 0 \end{aligned} \right\} \quad (\gamma)$$

which can be treated as a system of homogeneous equations in the unknowns A, B, C . This system has nontrivial solutions because all the coefficients A, B, C in (α) cannot simultaneously be zero. Therefore system (γ) has a nontrivial solution and hence

$$\begin{vmatrix} x_0 - x'_0 & y_0 - y'_0 & z_0 - z'_0 \\ m & n & p \\ m' & n' & p' \end{vmatrix} = 0 \quad (\delta)$$

This is the sought for condition for two lines to lie in one plane.

Now it can be easily demonstrated that if the two lines in question have a common point in plane ∞ , then m, n, p and m', n', p' are proportional. Indeed, since the common point of the two lines in question lies in plane ∞ , it has no nonhomogeneous coordinates. Therefore for no values of t and t' can (β) be reduced to the equations $x = x', y = y', z = z'$. If we assume that such a reduction is possible, we obtain a system of equations

$$\left. \begin{aligned} mt - m't' + (x_0 - x'_0) &= 0 \\ nt - n't' + (y_0 - y'_0) &= 0 \\ pt - p't' + (z_0 - z'_0) &= 0 \end{aligned} \right\} \quad (\epsilon)$$

in the numbers $t, -t', 1$. Since the determinant (δ) of this system vanishes, one of the equations is a linear combination of the other two. If, moreover, at least one of the three determinants $\begin{vmatrix} m & m' \\ n & n' \end{vmatrix}, \begin{vmatrix} n & n' \\ p & p' \end{vmatrix}, \begin{vmatrix} p & p' \\ m & m' \end{vmatrix}$ does not vanish, then (ϵ) possesses a solution, which,

as noted above, is impossible. Thus

$$\begin{vmatrix} m & m' \\ n & n' \end{vmatrix} = 0, \quad \begin{vmatrix} n & n' \\ p & p' \end{vmatrix} = 0, \quad \begin{vmatrix} p & p' \\ m & m' \end{vmatrix} = 0$$

implying

$$m \div n \div p = m' \div n' \div p'$$

as required.

Thus, all points in the space, without any exception have been assigned homogeneous coordinates. The basic properties of these coordinates are in complete analogy to the aforementioned properties of homogeneous coordinates on a line and on a plane.

The most striking feature of the homogeneous coordinates just introduced is that whatever the point M lying in the plane represented by the equation

$$Ax + By + Cz + D = 0 \quad (1)$$

in nonhomogeneous coordinates, the homogeneous coordinates x_1, x_2, x_3, x_4 of M always satisfy the relation

$$Ax_1 + Bx_2 + Cx_3 + Dx_4 = 0 \quad (2)$$

In fact, if M does not lie in plane ∞ , then $x = x_1/x_4$, $y = x_2/x_4$, $z = x_3/x_4$ ($x_4 \neq 0$) and (2) immediately follows from (1). If M lies in plane ∞ , then $x_4 = 0$ and $x_1 \div x_2 \div x_3 = m \div n \div p$, where m, n, p are parameters of equations of any line passing through point M . But, as seen earlier, A, B, C and m, n, p are connected by the relation

$$Am + Bn + Cp = 0$$

which yields

$$Ax_1 + Bx_2 + Cx_3 = 0$$

if one notes that $x_1 \div x_2 \div x_3 = m \div n \div p$. And this relation coincides with (2) if $x_4 = 0$.

The relation (2) connecting the homogeneous coordinates of points of the plane will be called the equation of this plane in homogeneous coordinates. Changing the notation of the coefficients in the equation of the plane, we see that this equation can be written in the form

$$u_1x_1 + u_2x_2 + u_3x_3 + u_4x_4 = 0$$

By the definition of homogeneous coordinates x_1, x_2, x_3, x_4 of any point M , at least one of these numbers is not zero. Multiplying the four coordinates by a suitable number, we see that the nonzero coordinate among them may be made equal to unity. For example, for homogeneous coordinates of point O one may take the four numbers 0, 0, 0, 1; of point ∞_x the numbers 1, 0, 0, 0; of point ∞_y the numbers 0, 1, 0, 0; of point ∞_z the numbers 0, 0, 1, 0;

and of point E the numbers 1, 1, 1, 1. In place of $\infty_x, \infty_y, \infty_z$ and O we shall henceforth write A_1, A_2, A_3, A_4 and call these points the *vertices of the coordinate tetrahedron*. Obviously, the system of homogeneous coordinates in projective space is determined by the choice of the coordinate tetrahedron $A_1(1, 0, 0, 0), A_2(0, 1, 0, 0), A_3(0, 0, 1, 0), A_4(0, 0, 0, 1)$ and the unit point $E(1, 1, 1, 1)$. These five points must be chosen subject to the sole condition that no four of them be coplanar.

The question as to how the homogeneous coordinates are affected by a change in the position of points determining the given system will be discussed in § 114, where formulas for transformation of homogeneous coordinates will be derived.

5.7. Projective correspondence between element of one-dimensional manifolds

§ 103. Fundamental for projective geometry is the notion of a projectivity. Suppose that between the points of two projective lines a and a' a one-to-one correspondence is established. If M is any point of a and M' is the corresponding point of a' , then M' will be called a function of M and the usual symbol of functional dependence will be used: $M' = f(M)$. It is clear that M may be treated as a function of M' , that is $M = \varphi(M')$; the functions $f(M)$ and $\varphi(M')$ may be appropriately called inverses of each other.

A one-to-one correspondence $M' = f(M)$ between points of two projective lines a and a' is called projective if to harmonically conjugate pairs of points M, N and P, Q of line a there always correspond harmonically conjugate pairs of points M', N' and P', Q' of line a' .

Prescribing the points $M' = f(M)$ of line a' which projectively correspond to points M of line a is referred to as the projectivity of line a into line a' . When a and a' coincide, we say that a projectivity of the line into itself is given.

An important particular case of projectivity of one line into another is provided by the central projection. Let a and a' be two lines lying in one plane α and S a point of α not on either of these lines. For the image of any point M of a take the point $M' = f(M)$ of a' which lies together with M on one line emanating from S . Such a mapping $M' = f(M)$ is a projectivity. In fact, according to Theorem 6 of § 86, if M_1, M_2 and M_3, M_4 are arbitrary points of a forming harmonically conjugate pairs, then the pairs of corresponding points M'_1, M'_2 and M'_3, M'_4 of line a' are also harmonically conjugate. And this is the characteristic property of a projectivity.

Thus a projectivity can be treated as a generalization of the central projection. Now we shall prove the following two simple theorems.

Theorem 14a. *If the mapping $M' = f(M)$ of line a into line a' is a projectivity, its inverse $M = \varphi(M')$ is also a projectivity.*

The proof is extremely simple. Indeed, suppose that the mapping $M = \varphi(M')$ is not a projectivity. Then on a' there are two pairs of harmonically conjugate points A', B' and C', D' to which there correspond on a two pairs of points $A = \varphi(A'), B = \varphi(B')$ and $C = \varphi(C'), D = \varphi(D')$ that are not harmonically conjugate. Denote by D^* the point of a which together with C forms a pair C, D^* that is harmonically conjugate to the pair A, B . Obviously, D and D^* are distinct.

Let $D^{*'} = f(D^*)$. Since the mapping $M' = f(M)$ is one-to-one, it follows that D' and $D^{*'}$ are also distinct. Because $M' = f(M)$ is a projectivity, the pairs A', B' and $C', D^{*'}$ are harmonically conjugate. Thus to three points A', B', C' one obtains distinct fourth harmonic points D' and $D^{*'}$, which is impossible. This contradiction proves the theorem.

Theorem 14b. *If $M' = f_1(M)$ is a projectivity of a line a into a line a' and $M'' = f_2(M')$ is a projectivity of line a' into another line a'' (the three lines may coincide, in particular), then the mapping $M'' = f_2 f_1(M)$ of a into a'' is also a projectivity.*

In other words, the mapping obtained by the successive application of two projectivities is also a projectivity.

The stated assertion is almost obvious. In fact, since both the mapping f_1 and f_2 preserve the harmonic conjugateness of pairs of points, the mapping obtained by their successive application also preserves the harmonic conjugateness of pairs of points and is hence a projectivity. ■

The properties possessed by the set of projectivities obtained in Theorem 14b is called the *group property* (the reader should consult § 19 again, where the group property was mentioned in connection with the set of motions).

In elementary geometry a system of points $M_1, M_2, \dots, M_n$ lying on a line a is said to be equivalent to a system of points $M'_1, M'_2, \dots, M'_n$ lying on the same line or on another line a' if there is a motion making the first system coincide with the second. Similarly, in projective geometry a system of points $M_1, M_2, \dots, M_n$ of a line a is said to be equivalent to the system of points $M'_1, M'_2, \dots, M'_n$ of a line a' if there is a projectivity of a into a' which maps every point M_i into point M'_i ($i = 1, 2, \dots, n$).

In particular, the system $M_1, M_2, \dots, M_n$ is equivalent (often the term *projectively equivalent* will also be used) to the system $M'_1, M'_2, \dots, M'_n$ if every point M_i is mapped into M'_i by a series of central projections of which the first projects line a into line a_1 , the second projects a_1 into $a_2, \dots$, and the last projects a_{n-1} into a' .

Theorems 14a and 14b imply that

(1) If a system of points lying on a line is projectively equivalent to the second system, then the second is projectively equivalent to the first.

(2) If two systems are projectively equivalent to a third, then they are projectively equivalent to each other.

Projectivity is a fundamental concept in projective geometry because it enables us to define the projective equivalence of two systems of points. In this sense it is comparable with the notion of congruent displacement (motion) in elementary geometry.

A careful study of projectivities will be made in the following subsections. Our first concern is to answer the question as to how a projectivity is determined uniquely. An answer to this question is provided by Staudt's theorem. This theorem will be formulated and proved below after three lemmas required for the proof have been established.

Lemma 1. *Let two pairs of points M, N and P, Q be given on a projective line a . In order for a third pair to exist on a which will be harmonically conjugate to both the pairs M, N and P, Q , it is necessary and sufficient that the pairs M, N and P, Q do not divide each other.*

Proof of necessity. Suppose that there is a pair of points X, Y which divides both the pairs M, N and P, Q harmonically. On line a introduce a projective (nonhomogeneous) coordinate system assigning the role of zero point to X and the role of point at infinity to Y and choosing the unit point arbitrarily. Then X (the zero point of the system) is the projective centre of the segment MN . Therefore, if x_1 and x_2 are coordinates of M and N , then $x_1 + x_2 = 0$. Accordingly, x_1 and x_2 , being of opposite signs, have the same absolute value x . Similarly, the coordinates of P and Q have a common absolute value y . We know that in the projective coordinate system on a deleted projective line the points and their corresponding coordinates are in the same order. Therefore, if $x < y$, the points M, N lie inside the segment PQ , while if $y < x$, the points P, Q lie inside the segment MN . But in either case, the pairs M, N and P, Q do not divide each other.

Proof of sufficiency. Assume that M, N and P, Q do not divide each other. We shall show that in this case there is always a pair X, Y which divides both the pairs P, Q and M, N harmonically. Since the pairs M, N and P, Q do not divide each other, both P and Q lie inside one of the two segments into which the points M, N divide the projective line. Inside this segment take any point E and on the projective line introduce a system of nonhomogeneous coordinates taking M as the zero point, N as the point at infinity and E as the unit point. Let p, q be the coordinates of P, Q . The numbers p, q are positive because of the unit point chosen. Assume further that $y = f(x)$ connects the coordinates of points X, Y which divide the pair P, Q harmonically. The function $y = f(x)$ is unde-

finer for $x = (p + q)/2$ because $(p + q)/2$ is the coordinate of the projective centre of the segment PQ , and this centre happens to be the fourth harmonic point to the three points P, Q, ∞ ; accordingly, for $x = (p + q)/2$ we find $y = f(x) = \infty$. For the remaining values of x the function $y = f(x)$ has definite numerical values and is continuous; the latter result follows from Theorem 11 and property (3) of projective coordinates mentioned in § 97. Notice now that $y \rightarrow \infty$ as $x \rightarrow (p + q)/2$ and $y \rightarrow -\infty$ as $x \rightarrow (p + q)/2$ and $x < (p + q)/2$; if $x = p$, then $y = p$ (see the Remark at the end of § 93) and, accordingly, $y > 0$. Suppose that $p < q$. Then, as x varies from p to $(p + q)/2$, the function $\varphi(x) = x + f(x)$, while remaining continuous, changes its sign from plus to minus. Hence there must be a value of x for which $x + f(x) = x + y = 0$. Let X, Y be points with just these coordinates x, y . They determine a segment XY with its projective centre at the zero point, M . In other words, the pair X, Y divides the pair M, N (recall that N has been taken as the point at infinity). Since, according to the definition of the function $y = f(x)$, the pair X, Y divides the pair P, Q harmonically, this is the desired pair of points. ■

Lemma 2. *The property of division of pairs of points is invariant under projectivities.*

This lemma is a direct consequence of the preceding lemma. In fact, let two pairs of points M, N and P, Q be given on a line a and suppose that these pairs do not divide each other. Then, by Lemma 1, there exists a pair of points X, Y which is harmonically conjugate to both the pairs M, N and P, Q . Under a projectivity of line a into another line a' the points M, N, P, Q and X, Y are mapped into points M', N', P', Q' and X', Y' in which the pair X', Y' is harmonically conjugate to both the pairs M', N' and P', Q' (this follows from the definition of a projectivity). But then Lemma 1 implies that the pairs M', N' and P', Q' do not divide each other. This shows that under a projectivity undividing pairs are mapped into undividing pairs, and then the dividing pairs are evidently mapped into dividing pairs.

In fact, if dividing pairs were mapped into undividing pairs, then under the inverse mapping (which is also a projectivity, see Theorem 14a) undividing pairs would become dividing pairs, which, as shown just now, is impossible. ■

The following lemma is of purely analytic nature.

Lemma 3. *Let $f(x)$ and $\varphi(x)$ be two functions with $f(x)$ monotone and $\varphi(x)$ continuous defined for any x , $-\infty < x < \infty$. Then, if $f(x)$ and $\varphi(x)$ coincide on an everywhere dense set of points of the number line, the two functions coincide identically.*

Let A be an everywhere dense set of points on the number line where the two functions $f(x)$ and $\varphi(x)$ assume the same values, and let x_0 be a point not belonging to A . We have to show that $f(x_0) =$

$= \varphi(x_0)$. Assume that $f(x_0) > \varphi(x_0)$. Since the set A is everywhere dense, we may choose two points x_1, x_2 in A such that for $x_1 < x_0 < x_2$ the difference $x_2 - x_1$ is as close to zero as one desires. Because $\varphi(x)$ is continuous, the quantities $\varphi(x_1)$ and $\varphi(x_2)$ differ so little from $\varphi(x_0)$, provided the difference $x_2 - x_1$ is sufficiently small, that, apart from the inequality $f(x_0) > \varphi(x_0)$, also $f(x_0) > \varphi(x_1)$ and $f(x_0) > \varphi(x_2)$. But since the functions $f(x)$ and $\varphi(x)$ assume the same values on A and x_1, x_2 are arbitrary points of A , it follows that $\varphi(x_1) = f(x_1)$ and $\varphi(x_2) = f(x_2)$. Hence we find that $f(x_1) < f(x_0)$, $f(x_2) < f(x_0)$ for $x_1 < x_0 < x_2$, and this contradicts the assumption that $f(x)$ is monotone. The assumption $f(x_0) < \varphi(x_0)$ similarly leads to a contradiction. ■

Now we are in a position to prove the fundamental theorem, due to Staudt, of projective geometry.

Theorem 15. *The projectivity between two lines is uniquely determined by three pairs of corresponding points.*

Proof. Let a and a' be two projective lines between which a projectivity has been established, assigning a point $M' = f(M)$ of a' to a point M of a . Assume that A, B, C are three distinct points of a and $A' = f(A), B' = f(B)$ and $C' = f(C)$ are their corresponding points lying on a' . We must show that there is no other projectivity $M' = \varphi(M)$ of a into a' , under which $A' = \varphi(A), B' = \varphi(B)$ and $C' = \varphi(C)$.

For the proof, we introduce a projective system of (nonhomogeneous) coordinates on line a by taking A as the zero point, B as the unit point and C as the point at infinity. Simultaneously projective coordinates are introduced on a' by choosing the points A', B', C' as the zero point, unit point and the point at infinity, respectively. Having introduced coordinate systems on a and a' , we see that every point M (except the point at infinity) of a may be characterized by its coordinate x and every point M' (except the point at infinity) of a' by its coordinate x' . And instead of examining the relation $M' = f(M)$, one may examine its arithmetical equivalent $x' = f(x)$, where x and x' are coordinates of the projectively corresponding points M and M' . The theorem will have been established if we are able to show that $x' = f(x)$ is a completely defined function. We shall now show that $f(x) \equiv x$.

One may easily discover the source of the identity $f(x) \equiv x$ by comparing the definitions of the projectivity and projective coordinates. First, since the points A, B, C of a and their corresponding points A', B', C' of a' , under the mapping $M' = f(M)$, have been chosen as the zero point, unit point and the point at infinity, we must have $f(0) = 0, f(1) = 1$ and $f(\infty) = \infty$. Next, the point D , labelled by 2 in the projective scale on line a , together with A forms a pair harmonically conjugate to the pair B, C , and, because projectivity preserves (by definition) harmonic conjugateness, the

point D must be mapped into a point D' such that the pair A', D' divides the pair B', C' , harmonically. Accordingly, point D' of a' , like D of a , has coordinate 2, that is, $f(2) = 2$. Arguing in the same manner, we find that $f(3) = 3$, $f(4) = 4$, etc., $f(-1) = -1$, $f(-2) = -2$, etc. Thus $f(n) = n$ for any integer n . The definition of a projectivity also implies that the projective centres of segments with integer endpoints on line a are mapped into projective centres of the corresponding segments with integer endpoints on line a' ; consequently, $f\left(\frac{n}{2}\right) = \frac{n}{2}$. In exactly the same way, the projective centres of segments $\left(\frac{n}{2}, \frac{n+1}{2}\right)$ on line a are mapped into projective centres of segments $\left(\frac{n}{2}, \frac{n+1}{2}\right)$ on line a' ; accordingly, $f\left(\frac{n}{2^2}\right) = \frac{n}{2^2}$, and so on. R

Thus $f(x) = x$ if x is a binary fraction. We have to show that $f(x) = x$ for any x . With this aim, note that $f(x)$, being defined for all x , $-\infty < x < +\infty$, is monotone. In fact, consider three points M_1, M_2, M_3 (all distinct from C) of line a and their corresponding points M'_1, M'_2, M'_3 of line a' , obtained through this mapping. Suppose that on the deleted projective line a the point M_2 lies between the points M_1 and M_3 ; this implies that the pair M_2, C divides the pair M_1, M_3 . But then Lemma 2 states that the pair M'_2, C' divides the pair M'_1, M'_3 . Accordingly, on the deleted line a' the point M'_2 lies between the points M'_1 and M'_3 . Thus the mapping $M' = f(M)$ in question either preserves the order of points or reverses it; depending on this, the function $x' = f(x)$ is either monotonically increasing or monotonically decreasing.

We have seen that $f(x) = x$ if x is a binary fraction. Accordingly, two functions $f(x)$ and $\varphi(x) = x$ assume equal values on an everywhere dense set of points on the number line (namely, on the set of binary fractions). Since the function $f(x)$ is monotone and $\varphi(x) = x$ is continuous, it follows from Lemma 3 that these functions coincide identically, that is, $x' = f(x) = \varphi(x) = x$ for any x .

Having established this fact, we have completed the proof of the theorem. In fact, as soon as three pairs of points A, A' ; B, B' ; C, C' are given which correspond under the projectivity $M' = f(M)$ of line a into line a' , to each point M there is assigned a point M' which has the same coordinate on a' as M has on a . This means that a projectivity is completely determined by three pairs of corresponding points. ■

An important corollary of the present theorem is

Theorem 16. *Under a projectivity other than the identity of a projective line into itself, the number of fixed points cannot exceed two.*

Proof. Let $M' = f(M)$ be a nonidentity projectivity that maps a projective line u into itself. Suppose that it has three fixed points

A, B, C , that is, there exist three points A, B, C which coincide with their corresponding points A', B', C' so that $A' = f(A) = A$, $B' = f(B) = B$, $C' = f(C) = C$. Apart from $M' = f(M)$, consider also the identity mapping which maps u into itself, that is, every point M coincides with its corresponding point M' : $M' \equiv M$. All the points of u , including A, B, C are fixed points of the identity mapping. Thus both the mappings $M' = f(M)$ and the identity map the points A, B, C into the same points A, B, C . Therefore both these mappings have three pairs of corresponding points. Since each of the mappings is a projectivity (the mapping $M' = f(M)$ by hypothesis and the mapping $M' \equiv M$ in an obvious manner), the preceding theorem implies that the two mappings are not different from each other. In other words, the mapping $M' = f(M)$ must be an identity, contrary to the hypothesis of the theorem. Thus the assumption that $M' = f(M)$ has three fixed points leads to a contradiction. ■

The same result can be alternatively formulated in the following

Theorem 17. *If a projectivity of a line into itself has three fixed points, then all the points of the line are fixed points, that is, the mapping must be the identity.*

§ 104. We shall call the following *one-dimensional projective manifolds*:

- (1) the set of points of a projective line;
- (2) the set of rays of a plane pencil, that is, the set of lines lying in one plane and passing through a point of the plane, called the centre of the pencil;
- (3) the set of planes passing through a line in space (such a set of planes is called a *pencil* and the line through which the planes pass the *axis of the pencil*).

The concept of a projective correspondence introduced above for projective lines can be extended in a natural way to the case of arbitrary one-dimensional manifolds.

Let Π and Π' be any two one-dimensional manifolds. Suppose that a one-to-one correspondence is set up between the elements of the two manifolds such that to any element x of Π there is assigned an element $x' = f(x)$ of Π' . The correspondence $x' = f(x)$ will be called a *projectivity* if it assigns harmonically conjugate pairs of elements x'_1, x'_2 and x'_3, x'_4 of Π' to harmonically conjugate pairs of elements x_1, x_2 and x_3, x_4 of Π .

The following theorem generalizes Theorem 15 to the case of a projectivity between arbitrary one-dimensional manifolds.

Theorem 18. *The projectivity between two one-dimensional manifolds is uniquely determined by three pairs of corresponding elements.*

Proof. Let two one-dimensional manifolds Π and Π' be given and let a projectivity between them be given that assigns an element $x' = f(x)$ of Π' to an element x of Π . Suppose that a, b, c are three

distinct elements of Π and a', b', c' the corresponding elements of Π' . We have to show that there is no projectivity other than the given $x' = f(x)$, between Π and Π' which also assigns a', b', c' to a, b, c .

In order to simplify the presentation, consider a definite case, assuming, for example, that Π and Π' are plane pencils of rays. In Π take any line u not passing through the centre of this pencil, and, similarly, take a line u' in Π' . Let X be the point where the ray x of Π intersects u , and X' the point where the ray $x' = f(x)$ of Π' intersects u' . Let us examine the correspondence between u and u' through which point X' is assigned to point X , writing $X' = F(X)$. It may easily be seen that the correspondence $X' = F(X)$ is a projectivity. This follows immediately from the definition of a projective correspondence between pencils (which has been formulated above for any one-dimensional manifolds) and the proposition formulated at the end of § 86 stating that a projection and a section preserve harmonic conjugateness.

Thus the projective correspondence $x' = f(x)$ between the pencils Π and Π' induces a projective correspondence $X' = F(X)$ between the lines u and u' . And it is clear that distinct correspondences $x' = f(x)$ and $x' = \varphi(x)$ between Π and Π' induce distinct correspondences $X' = F(X)$ and $X' = \Phi(X)$ between u and u' . Let A, B, C be the points of intersection of line u with the rays a, b, c and A', B', C' the points of intersection of line u' with the rays a', b', c' . If, apart from the projective correspondence $x' = f(x)$, there were another correspondence $x' = \varphi(x)$ between Π and Π' which associates, like the first, the rays a, b, c with the rays a', b', c' , then there would be distinct projective correspondences $X' = F(X)$ and $X' = \Phi(X)$ between u and u' so that both $X' = F(X)$ and $X' = \Phi(X)$ would map the points A, B, C into the points A', B', C' . But this would contradict Theorem 15. Accordingly, there is no projective correspondence other than $x' = f(x)$ between the pencils Π and Π' that maps a, b, c into a', b', c' . Thus the projective correspondence between the pencils is defined by three pairs of corresponding rays.

If Π and Π' are some other one-dimensional manifolds, then by means of a section the projective correspondence between Π and Π' may be reduced to the one between lines and thus in all cases Theorem 18 may be obtained from Theorem 15. ■

The following is an obvious corollary of Theorem 18.

Theorem 19. *A projectivity, other than the identity, of any one-dimensional manifold into itself cannot have more than two fixed points.*

This theorem is a generalization of Theorem 16 which concerned fixed points of a projectivity of a line into itself.

§ 105. We mention one more proposition that will be required later.

Theorem 20. *Let two one-dimensional projective manifolds Π and Π' be given. Suppose that an element $x' = f(x)$ of Π' is assigned to each element x of Π , where distinct elements $x'_1 = f(x_1)$ and $x'_2 = f(x_2)$ are assigned to distinct elements x_1 and x_2 . If harmonically conjugate pairs of elements of Π' correspond to harmonically conjugate pairs of elements of Π , then $x' = f(x)$ is a one-to-one mapping of Π into Π' .*

Proof. It suffices to examine the case when Π and Π' are lines because all other cases are reduced to this case by means of a section, as was done in the proof of Theorem 18.

Suppose that Π and Π' are lines, say u and u' , and suppose that every point M of u is mapped into a point $M' = f(M)$ of u' ; moreover, distinct points of u are mapped into distinct points of u' and harmonically conjugate pairs of points of u are mapped into harmonically conjugate pairs of points of u' . We have to show that $M' = f(M)$ is a one-to-one mapping of u into u' , that is, that every point of u' is the image of some point of u .

It is easy to see that this assertion immediately follows from arguments used in proving Theorem 15. Indeed, take three points on u and label two of them by the numbers 0 and 1 and the third by the symbol ∞ . Their images on u' will also be labeled by 0, 1 and ∞ , respectively. Next, introduce nonhomogeneous projective coordinate systems on both u and u' by means of the points 0, 1 and ∞ . Then the symbolic relation $M' = f(M)$ may be replaced by the arithmetical relation $x' = f(x)$ between the coordinates of points M and M' .

The theorem will have been proved if we establish that the function $x' = f(x)$ assumes all the values between $-\infty$ and $+\infty$ as x varies from $-\infty$ to $+\infty$. And, repeating the arguments used in proving Theorem 15, we would have concluded that $f(x) \equiv x$, from which the desired result would follow.

However, in this argument there is a snag. Namely, Theorem 15 uses Lemma 2 concerning projectivity of one line into another. The mapping now under discussion preserves, like projectivity, harmonic conjugateness of pairs of points, but, in contrast to projectivity, is not assumed to be one-to-one beforehand.

Therefore, before applying Lemma 2 to the present case, the proof of this lemma must be freed from the one-to-oneness of the mapping. Recall that the proof there consisted of two parts. The first part contained the proof of the fact that the projectivity $M' = f(M)$ of u into u' maps the undividing pairs of points of u into undividing pairs of points of u' ; this part did not require that the mapping be one-to-one. The second part consisted in showing that the dividing pairs of u are mapped into dividing pairs of u' . This fact was proved by using the mapping inverse of $M' = f(M)$, and the existence of an inverse amounts to the one-to-oneness of the mapping $M' = f(M)$. Therefore this part of the proof of Lemma 2 must be modified. This

modification does not require much work. Having shown that $M' = f(M)$ maps undividing pairs into similar pairs, the proof of the fact that dividing pairs are mapped into dividing pairs can be carried out by the "method of contradiction" as follows. Suppose that to two dividing pairs of points A, B and C, D on u there correspond on u' undividing pairs of points A', B' and C', D' . Since the pairs A, B and C, D divide each other, Axiom II.3 implies that the pairs A, C and B, D and the pairs A, D and B, C must not divide each other. On the other hand, since the pairs A', B' and C', D' do not divide each other, the same Axiom II.3 implies that either the pairs A', C' and B', D' or the pairs A', D' and C', B' are dividing pairs. Thus our hypothesis implies that there are necessarily undividing pairs on u that are mapped into dividing pairs on u' , which is contrary to the initial premise of the argument. ■

5.8. Projective correspondences between two- and three-dimensional manifolds

§ 106. We shall now define the projectivity of two- and three-dimensional images.

First consider the two-dimensional case. Suppose that between the points of two planes α and α' a one-to-one correspondence has been set up which assigns a point $M' = f(M)$ of α' to a point M of α .

This correspondence is called *projective* if to points lying on one line in α there correspond points also lying on one line in α' .

When points $M' = f(M)$ of α' projectively correspond to points M of α , we also speak of a *projectivity of plane α into plane α'* . If α and α' coincide, we say that the projectivity maps α into itself.

By the definition of a projectivity of α into α' , points of every line a in α have images which lie in α' on some line a' . This line a' will be said to correspond to line a under the given mapping.

The definition of a projective correspondence requires that points lying on one line be mapped into points also lying on one line. However, the definition says nothing about the points not lying on one line and the possibility of such points being mapped into points of one line is not rejected beforehand. But, it will be shown later that this case cannot occur, that is, if the image points lie on one line, their preimages also lie on one line. In other words, we shall show that the inverse mapping of a projectivity is also a projectivity (Theorem 23a). At the same time, it will also be established that under a projectivity, the correspondence between lines, just as in the case of points, is one-to-one.

An important particular case of the projectivity of a plane into another plane is the central projection.

If the points of a plane α are projected from any centre to another plane α' (as onto a screen), every point M of α is mapped into a

point $M' = f(M)$ of α' . The correspondence $M' = f(M)$ is a projectivity because any line in α is mapped into a line in α' .

Let us prove the following

Theorem 21. *If a projectivity maps a plane α into a plane α' , then the set of harmonic elements of α is mapped into a set of harmonic elements of α' .*

Proof. (1) Let a be any line in α , a' its corresponding line in α' , and let A, B and C, D be arbitrarily chosen harmonically conjugate pairs of points of a . We have to show that the pairs of points A', B' and C', D' of line a' , which are maps of A, B, C, D are also harmonically conjugate. First of all, note that there is a point in α outside a whose image does not lie on a' . In fact, if all the points of α not lying on a were mapped into points lying on a' , then a set of points of α would have been mapped into a set of points of α' not lying on a' (because the projectivity of α into α' is assumed to be one-to-one), but this is contrary to the definition of the projectivity (which preserves the rectilinear character of the location of points). Let R be a point in α not lying on a whose image R' in α' does not lie on a' . Since the pairs A, B and C, D are harmonically conjugate, a four-point T may be constructed in α whose diagonal points are A, B and the pairs of opposite sides pass through C, D ; moreover, R may be taken as one of the vertices of T (see § 86). Since the image R' of R is not on a' , it follows that no three images of the vertices of T are colinear. Therefore the image of a four-point T is some four-point T' .

Evidently, the points A', B' are diagonal points of T' and two opposite sides of T' pass through C', D' , implying that the pairs A', B' and C', D' are harmonically conjugate.

(2) Let P be any point in α and P' its image in α' , and let a, b and c, d be arbitrarily chosen harmonically conjugate pairs of rays of a pencil in α with centre P . We have to show that in the pencil with centre P' the rays a', b' and c', d' corresponding to the rays a, b, c, d are also harmonically conjugate. This immediately follows from the preceding result. First of all note that in α there is a line not passing through P whose image does not pass through P' . In fact, take a point Q , distinct from P , in α and denote its image in α' by Q' . As shown above, there is a point R in α that does not lie on the line PQ . Evidently, the line QR is just the line which does not pass through P and whose image does not pass through P' . Denote the line QR by t and its image by t' . Let A, B, C, D be the points where the rays a, b, c, d intersect the line t and, similarly, let A', B', C', D' be the points where the rays a', b', c', d' intersect t' . Clearly, A', B', C', D' are images of A, B, C, D . Since the pairs of rays a, b and c, d are harmonically conjugate, it follows from the proposition mentioned at the end of § 86 that the pairs of points A, B and C, D are harmonically conjugate. This, in view of the first part of the present

proof, implies that the pairs A', B' and C', D' , being images of the points A, B, C, D , are also harmonically conjugate. But, since the rays a', b', c', d' pass through the points A', B', C', D' , respectively, the same proposition at the end of § 86 implies that the pairs of rays a', b' and c', d' are harmonically conjugate. The theorem is completely proved. ■

Theorems 20 and 21 imply

Theorem 22. *If a plane α is mapped by a projectivity into plane α' , then (i) the set of points of every line a in α is mapped in a one-to-one manner into the set of points of the corresponding line a' in α' . and (ii) the set of rays of any pencil in α with centre P is mapped in a one-to-one manner into the set of rays of the pencil whose centre P is a point in α' corresponding to the point P .*

From this theorem the following result can be easily derived.

Theorem 23a. *If $M' = f(M)$ is a projectivity of a plane α into plane α' , then the inverse mapping $M = \varphi(M')$ of α' into α is also a projectivity.*

Proof. Let a' be any line in α' . On this line take any two points A', B' ; to these points there correspond points $A = \varphi(A'), B = \varphi(B')$ in α . Denote by a the line determined by A and B . Since $M' = f(M)$ is a projectivity, it maps all the points of a into points of a' . According to Theorem 22, this mapping of a into a' is one-to-one, that is, images of points of a "fill" the line a' . In other words, every point of a' is the image of some point of a , which means that if M' lies on a' , the point $M = \varphi(M')$ lies on a . Thus $M = \varphi(M')$ maps points of α' lying on any one line into points that lie on one line in α , which is characteristic of a projectivity. ■

It is interesting to note that the following theorem holds, which at first sight appears surprising.

Suppose that a set of all points of a plane α is mapped in a one-to-one manner into a set of points G' of plane α' . If the points of α lying on any one line are mapped into points of α' also lying on one line, then only the following two cases are possible: either (a) G' lies totally on any one line in α' or (b) G' coincides with the entire plane α' (in this case the projectivity maps α into entire plane α').

Proof. The first case is obtained if any point set G' having the power of the continuum is taken on any line in α' and α is mapped in an arbitrary way in a one-to-one manner into G' .

Suppose now that G' contains points of α' that do not lie on one line. Then the set of harmonically conjugate elements of α is mapped into a set of harmonically conjugate elements of α' (the proof of this fact is analogous to that of Theorem 21).

This together with Theorems 20, 21 implies that (1) the set of points of every line a in α is mapped in a one-to-one manner into the set of points of the corresponding line a' in α' , and (2) the set of rays

of any pencil in α with centre P is mapped in a one-to-one manner into the set of rays of a pencil whose centre P' is a point of α' corresponding to P .

Take any point P in α and let its image in α' be P' . Let M' be a totally arbitrary point of α' and let a' be the line joining M' with P' . In accordance with the earlier result, the line a' , as a ray of a pencil in α' with centre P' , must correspond to some line a belonging to a pencil with centre P in α ; moreover, the correspondence between the points of lines a and a' must be one-to-one. Therefore the point M' of a' must correspond to the point M of a , that is, to some point of α . Thus the images of points of α necessarily fill the entire plane α' . ■

Next we note the following obvious theorem.

Theorem 23b. *If $M' = f_1(M)$ is a projectivity of a plane α into a plane α' and $M'' = f_2(M')$ is a projectivity of plane α' into a plane α'' (in particular, all three planes may coincide with each other), then $M'' = f_2(f_1(M))$ mapping α into α'' is also a projectivity.*

In other words, a mapping obtained by applying two projectivities successively is also a projectivity.

This assertion is obvious. In fact, since both the mapping f_1 and f_2 preserve rectilinear location of points, the mapping obtained by applying them successively has the same property and hence is a projectivity.

The property expressed by Theorem 23b which a set of projectivities possesses is known as the group property.

We shall say that a figure Σ lying in a plane α is *projectively equivalent* to a figure Σ' lying in the same plane or in another plane α' if there is a projectivity of α into α' that maps Σ into Σ' .

In particular, figure Σ is projectively equivalent to figure Σ' if a chain of central projections of plane α into α_1 , of plane α_1 into plane α_2 , . . . , of plane α_{n-1} into plane α' results in Σ being projectively mapped into Σ' .

Theorems 23a and 23b imply that

- (i) if one figure is projectively equivalent to another, then the second is projectively equivalent to the first;
- (ii) if two figures are projectively equivalent to a third, then the two figures are projectively equivalent among themselves.

Because of the above association between figures, projectivities play the same role in projective geometry as congruent displacements of figures (that is, motions) do in elementary geometry. For some time projectivities of planes will be the subject of our study.

Theorem 24. *If a plane α is mapped by a projectivity into a plane α' , then (a) every line a in α is projectively mapped into the corresponding line a' in α' , and (b) every pencil of rays in α is projectively mapped into the corresponding pencil of rays in α' .*

To see that this theorem is indeed valid, it suffices to combine Theorems 21 and 22 and the definition of a projectivity for one-dimensional manifolds.

Now we are in a position to prove the following important theorem, which may be regarded as the extension of Theorem 15 to the case of two-dimensional manifolds.

Theorem 25. *A projectivity of plane α into plane α' is uniquely determined by four pairs of corresponding points subject to the condition that no three of the four points in α are colinear.*

Proof. Let a projectivity $M' = f(M)$ of plane α into plane α' be given. Suppose that A, B, C, D are four points in α , no three of which are colinear, and A', B', C', D' are the corresponding points in α' (the definition of a projectivity and Theorem 23a imply that among the points A', B', C', D' also no three are colinear). We have to show that there is no projectivity, other than $M' = f(M)$, of α into α' which would also map A, B, C, D into A', B', C', D' .

Take any point M in α and denote by m the line AM . This line is one of the rays constituting the pencil with centre A . Whatever the projectivity of α into α' which maps A, B, C, D into A', B', C', D' , a mapping of the pencil with centre A into a pencil with centre A' is a projectivity (see Theorem 24). Further, no matter how many projectivities of α into α' there are that map A, B, C, D into A', B', C', D' , all of them define a common projectivity of the pencil with centre A into the pencil with centre A' . In fact, each of them maps the rays AB, AC and AD of the first pencil into the rays $A'B', A'C'$ and $A'D'$ of the second; moreover, because of the choice of points A, B, C, D the rays AB, AC and AD (as well as $A'B', A'C'$ and $A'D'$) are distinct. But, according to Theorem 18, a projectivity of one-dimensional manifolds (in particular, of pencils) is uniquely determined by three pairs of corresponding elements. Therefore under all possible projectivities of α into α' that map A, B, C, D into A', B', C', D' , the line m is mapped into a completely defined line m' which is associated with m under the projective correspondence between the pencils in question. Accordingly, under all possible projectivities of α into α' that map A, B, C, D into A', B', C', D' , the point M is mapped into a completely defined line passing through A' . Similarly, by considering the pencil of rays in α with centre B and its map into α' one may establish that no matter how many projectivities of α into α' there are which map A, B, C, D into A', B', C', D' , the point M is mapped into a completely defined line passing through B' in α' . The intersection of these lines determines the image of M in α' , and this image is the same for all the projectivities of α into α' that map A, B, C, D into A', B', C', D' . And since M is an arbitrary point, it follows from the above arguments that there is no other projectivity apart from the given $M' = f(M)$ or α into α' that would map A, B, C, D into A', B', C', D' . ■

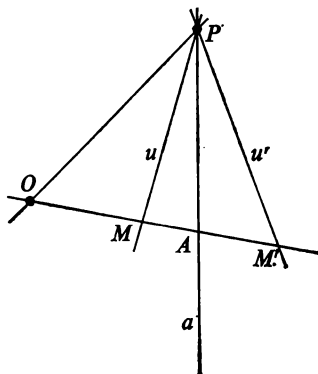


Figure 115

Theorem 25 implies

Theorem 26. *Under a projectivity, other than an identity, of a plane into itself there cannot be four fixed points of which three are collinear.*

Indeed, let $M' = f(M)$ be a projectivity, other than the identity, of a plane α into itself. Suppose that there are four points A, B, C, D , no three of which are collinear, which remain fixed under the mapping $M' = f(M)$. Besides $M' = f(M)$, consider the identity mapping $M' \equiv M$. It is evidently a projectivity. Further, under the mapping $M' \equiv M$ the points A, B, C, D (as do all points of the plane) remain fixed. Thus both $M' = f(M)$ and $M' \equiv M$ map A, B, C, D into the same points A, B, C, D . These mappings have therefore four common pairs of corresponding elements situated as in Theorem 25, and according to this theorem these mappings do not differ from each other. In other words, our assumption implies that the mapping $M' = f(M)$ must be an identity, contrary to the hypothesis of the theorem. ■

Remark. The condition imposed on the location of points in Theorems 25 and 26 is important. In order to see this, consider the so-called *harmonic mapping*.

Let O be any point in plane α and a any line in this plane not through O . Let M be any point of α and A the point where line OM intersects line a (Fig. 115). The point M' which together with point M divides the pair O, A harmonically will be treated as the corresponding point of M under the harmonic mapping of α into itself, the point O will be called the *centre of the mapping* and the line a its *axis*. Point M' will also be denoted by $M' = H(M)$.

It may be easily established that the harmonic mapping $M' = H(M)$ is a projectivity. In fact, let u be any line in α , P its point of intersection with axis a and u' the line passing through P that together with u divides the pair a, PO harmonically. Evidently, if M

is displaced along u , its corresponding point $M' = H(M)$ is displaced along u' . Thus under the mapping $M' = H(M)$ a line is mapped into a line, and this is the characteristic feature of a projectivity.

Further, according to the Remark made at the end of § 93, if point M coincides with a point of axis a , then the corresponding point $M' = H(M)$ coincides with the same point of the axis. Analogously, if M coincides with centre O , then $M' = H(M)$ also coincides with O . Accordingly, the centre O and all the points of axis a are fixed points of the mapping $M' = H(M)$. We find that the projectivity $M' = H(M)$ of α into itself has infinitely many fixed points (but all of them, except O , lie on one line). Thus, if the projectivity of a plane into itself has four fixed points, one may conclude that this mapping is an identity only when it is known that the location of the fixed points satisfy the condition of Theorem 26.

§ 107. We shall be frequently using the term *two-dimensional projective manifold* below meaning thereby either the projective plane or the so-called bundle. A *bundle* is the set of lines and planes in the projective space which pass through a point (*centre of the bundle*) of this space.

A projectivity may be defined for any two-dimensional manifolds. In order to formulate this definition in a convenient way, by an *element of the first kind* of a two-dimensional manifold Π we shall mean a point if Π is a plane or a line if Π is a bundle; by an *element of the second kind* of a two-dimensional manifold Π we shall mean a line if Π is a plane or a plane if Π is a bundle.

Suppose that between the elements of the first kind of two manifolds Π and Π' a one-to-one correspondence is set up that assigns an element $x' = f(x)$ of Π' to any element x of Π . The correspondence (or mapping) $x' = f(x)$ is known as a *projectivity* if to any set of elements of the first kind of Π belonging to one element of the second kind of this manifold there corresponds in Π' a set of elements of the first kind also belonging to one element of the second kind.

All the theorems established in the preceding subsection can be generalized in a natural way to arbitrary two-dimensional manifolds. In order to formulate these generalizations, in all the theorems of § 106 the words "point" and "line" should be replaced by the expressions "element of the first kind" and "element of the second kind". To justify this change, it suffices to observe that the intersection of elements of the first and second kind of the bundle with a plane results in elements of the first and second kind of this plane, respectively. Thus the investigation of projectivities of two-dimensional manifolds can be reduced to that of projectivities of planes, and this was carried out above.

§ 108. Let us now examine the projectivities of *three-dimensional projective manifolds*; by three-dimensional projective manifolds we mean projective spaces.

Consider two projective spaces Π and Π' (both the symbols Π and Π' denote certain sets of objects called points, lines and planes among which the incidence and order relationships are defined in conformity with the conditions of projective axioms). Suppose that between the points of Π and Π' a one-to-one correspondence has been established that assigns a point $M' = f(M)$ of Π' to any point M of Π . The one-to-one correspondence $M' = f(M)$ between the points of Π and Π' , is called a *projectivity* if the points corresponding to points of any one plane in Π also lie in one plane in Π' .

Prescribing points $M' = f(M)$ in Π' which projectively correspond to points M of Π is also referred to as the *projectivity of space Π into space Π'* . If Π and Π' coincide, one says that a projectivity of space into itself is defined.

The basic properties of projectivities of three-dimensional manifolds are a natural generalization of corresponding properties of projectivities of two-dimensional manifolds, and they may be established by using analogous arguments as was done in § 107. Therefore we shall confine ourselves to formulating the main theorems on projectivities in the three-dimensional case, omitting their proofs.

Theorem 27. *If a projectivity maps a space Π into space Π' , then a set of harmonic elements of Π is mapped into a set of harmonic elements of Π' .*

Theorem 28. *For a projectivity of space Π into space Π' the following holds:*

(1) *Every one-dimensional manifold of Π is mapped projectively in a one-to-one manner into the corresponding one-dimensional manifold of Π' . In particular, every line in Π is mapped projectively in a one-to-one manner into the corresponding line in Π' .*

(2) *Every two-dimensional manifold of Π is mapped projectively in a one-to-one manner into the corresponding two-dimensional manifold of Π' . In particular, every plane in Π is mapped projectively in a one-to-one manner into the corresponding plane of Π' .*

Theorem 29a. *If $M' = f(M)$ is a projectivity of space Π into space Π' , its inverse $M = \varphi(M')$ that maps Π' into Π is also a projectivity.*

Theorem 29b. *If $M' = f_1(M)$ is a projectivity of space Π into space Π' and $M'' = f_2(M')$ is a projectivity of space Π' into space Π'' , then the mapping $M'' = f_2(f_1(M))$ that transforms Π into Π'' is also a projectivity, that is, a set of projectivities of spaces possesses the group property.*

Theorem 30. *A projectivity of space Π into space Π' is uniquely determined by five pairs of corresponding points subject to the condition that no four of these five points of Π are coplanar.*

Theorem 30 immediately implies

Theorem 31. *A projectivity, other than an identity, of a space into itself cannot have five fixed points of which four are coplanar.*

The restriction imposed on the location of points in this theorem is important. One may see this by generalizing to a space the concept of harmonic mapping, which was presented for the plane case at the end of § 106.

Besides the above theorems, we also mention the following theorem.

Suppose that the set of all points of a projective space Π is mapped in a one-to-one manner into a set of points G' of the projective space Π' . If all the points of space Π lying in one plane are mapped into points of Π' also lying in one plane, then only the following two cases are possible: either (i) G' lies entirely in one plane in Π' or (ii) G' coincides with the entire space Π' (in this case the given mapping is a projectivity of space Π into the entire space Π').

This theorem is a natural generalization of the theorem concerning the mapping of a plane, which was stated and proved in § 106 after Theorem 23a.

The concept of the projective equivalence is defined for spatial bodies along the same lines as was done for figures of one- and two-dimensions. A body T of space Π is said to be *projectively equivalent* to body T' of the same space or of another space Π' if there is a projectivity of Π into Π' under which T is mapped into T' .

That this projective equivalence is both reflexive and transitive follows directly from Theorems 29a and 29b.

5.9. Analytic representation for projectivities.

Involution

§ 109. Our aim now will be to obtain formulas which connect projective coordinates of points that correspond to each other under a projectivity.

First we shall examine projectivities of a plane into a plane and a space into a space and then examine the one-dimensional manifolds, namely the projectivities of lines.

Let α , α' be two (not necessarily distinct) planes. In both of these planes introduce a system of homogeneous projective coordinates (if α and α' coincide, the coordinate systems introduced on them also may coincide). Next, we define a special mapping of points of α into points of α' ; namely, choosing the numbers c_{11} , c_{12} , c_{13} , c_{21} , c_{22} , c_{23} , c_{31} , c_{32} , c_{33} , we treat a point M' (x'_1 , x'_2 , x'_3) of α' as corresponding to the point M (x_1 , x_2 , x_3) of α under the mapping in question if the coordinates of these points satisfy the equations

$$\left. \begin{aligned} \rho'x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho'x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho'x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3 \end{aligned} \right\} \quad (1)$$

where ρ' is any number $\neq 0$. Such a mapping will be termed a *linear mapping* and we shall write $M' = L(M)$.

Recall that according to the basic property of homogeneous coordinates the choice of ρ' does not affect the location of point M' with coordinates $\rho'x'_1, \rho'x'_2, \rho'x'_3$ (see § 101). Therefore every point M can have only one image point M' . The quantities c_{ik} defining a linear mapping will be written in the form of a matrix

$$C = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix} \quad (2)$$

which will be called the *matrix of the linear mapping*, and the quantities c_{ik} will be called the *coefficients of the mapping*, while the determinant

$$\Delta = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} \quad (3)$$

will be called the *determinant of the mapping*.

It is important to notice that if $\Delta = 0$, not every point of α has an image. Indeed, if $\Delta = 0$, the system

$$\left. \begin{aligned} c_{11}x_1 + c_{12}x_2 + c_{13}x_3 &= 0 \\ c_{21}x_1 + c_{22}x_2 + c_{23}x_3 &= 0 \\ c_{31}x_1 + c_{32}x_2 + c_{33}x_3 &= 0 \end{aligned} \right\} \quad (*)$$

has a solution x_1^0, x_2^0, x_3^0 , which are not zero simultaneously; the point $M(x_1^0, x_2^0, x_3^0)$ has no image in α' because in this case (1) lead to $x'_1 = 0, x'_2 = 0, x'_3 = 0$, which is impossible since x'_1, x'_2, x'_3 are the homogeneous coordinates (see § 101). Thus the domain of definition of the mapping $M' = L(M)$ may not be the entire plane α but α with a certain point set $\alpha_0(M)$ excluded. It may be easily demonstrated that if $\Delta = 0$, the set $\alpha_0(M)$ is either a point or a line or the entire plane α . Namely, if the system (*) contains two separate (that is, linearly independent) equations, then (*) determines the ratio of the unknowns x_1, x_2, x_3 ; in this case $\alpha_0(M)$ consists of one point. If the system (*) contains only one separate equation, then $\alpha_0(M)$ is obviously a line (represented by this equation), and, finally, if the system (*) does not contain any separate equation (that is, if all $c_{ik} = 0$), then $\alpha_0(M)$ is the entire plane α .

In other words, $\alpha_0(M)$ is a point, a line or the entire plane corresponding to Rank $C = 2$, Rank $C = 1$ or Rank $C = 0$. The last case will be omitted from our discussion.

The following theorems hold for linear mappings.

Theorem 32. *If the determinant of a linear mapping $M' = L(M)$ transforming points of a plane α into points of plane α' does not vanish, then $M' = L(M)$ is a one-to-one mapping of α into α' .*

Proof. Consider a linear mapping $M' = L(M)$ defined by formulas (1) with determinant $\Delta \neq 0$. Then

(a) Every point M of α has an image in α' . Indeed, taking any point $M(x_1, x_2, x_3)$, we see that formulas (1) always define the numbers $\rho'x'_1, \rho'x'_2, \rho'x'_3$; these numbers cannot be all zero because with $\Delta \neq 0$ the equations

$$c_{11}x_1 + c_{12}x_2 + c_{13}x_3 = 0$$

$$c_{21}x_1 + c_{22}x_2 + c_{23}x_3 = 0$$

$$c_{31}x_1 + c_{32}x_2 + c_{33}x_3 = 0$$

would lead to $x_1 = x_2 = x_3 = 0$, which is impossible. But in homogeneous coordinates in α' the three numbers $\rho'x'_1, \rho'x'_2, \rho'x'_3$, if not all zero, determine a point M' that is the desired image of M .

(b) Every point M' of α' is the image of only one point M of α .

In fact, if $\Delta \neq 0$, equations (1) always uniquely determine the corresponding values x_1, x_2, x_3 for given x'_1, x'_2, x'_3 and ρ' ; if x'_1, x'_2, x'_3 are not all zero, then x_1, x_2, x_3 are also not all zero. Thus equations (1) determine homogeneous coordinates of some point M for the given homogeneous coordinates of M' . Furthermore, for different values of ρ' equations (1) determine different x_1, x_2, x_3 , but the ratio $x_1 \div x_2 \div x_3$ does not change with ρ' . Hence M is uniquely determined by a given M' . This completes the proof of the theorem. ■

If we set $\rho' = 1/\rho$ in (1) and express $\rho x_1, \rho x_2, \rho x_3$ from these equations, assuming $\Delta \neq 0$, we obtain the equations

$$\left. \begin{aligned} \rho x_1 &= c'_{11}x'_1 + c'_{12}x'_2 + c'_{13}x'_3 \\ \rho x_2 &= c'_{21}x'_1 + c'_{22}x'_2 + c'_{23}x'_3 \\ \rho x_3 &= c'_{31}x'_1 + c'_{32}x'_2 + c'_{33}x'_3 \end{aligned} \right\} \quad (4)$$

defining a mapping of α' into α , which is the inverse of the given mapping of α into α' . Obviously, the inverse mapping, like the one given, is linear. The coefficients c'_{ih} are expressed in terms of the coefficients c_{ih} according to the known formulas of algebra. They constitute the matrix

$$C' = \begin{pmatrix} c'_{11} & c'_{12} & c'_{13} \\ c'_{21} & c'_{22} & c'_{23} \\ c'_{31} & c'_{32} & c'_{33} \end{pmatrix} \quad (5)$$

which is the inverse of the matrix C of the given mapping, that is,

$$CC' = I$$

where

$$I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is the identity matrix. In this case, as is known,

$$\Delta\Delta' = 1$$

This, in particular, implies that $\Delta' \neq 0$.

Theorem 33. *If the determinant of a linear mapping $M' = L(M)$ that transforms points of a plane α into plane α' vanishes, then all the images of points of α lie on one line in α' . (In this case the mapping $M' = L(M)$ is not one-to-one.)*

Indeed, if the value of the determinant of a linear mapping defined according to formulas (1) is zero, then there exist three numbers u_1 , u_2 , u_3 , at least one of which is not zero, satisfying the equations

$$\left. \begin{aligned} c_{11}u_1 + c_{21}u_2 + c_{31}u_3 &= 0 \\ c_{12}u_1 + c_{22}u_2 + c_{32}u_3 &= 0 \\ c_{13}u_1 + c_{23}u_2 + c_{33}u_3 &= 0 \end{aligned} \right\} \quad (6)$$

Multiplying the first equation in (1) by u_1 , the second by u_2 and the third by u_3 and summing them and using (6), we obtain

$$u_1x'_1 + u_2x'_2 + u_3x'_3 = 0$$

for any x_1, x_2, x_3 . Thus, no matter how M is chosen, the coordinates of $M' = L(M)$ satisfy the equation of some line. This proves the theorem.* ■

Theorem 34. *Given a linear mapping $M' = L(M)$ of a plane α into a plane α' , the images of points of any one line in α also lie on one line in α' .*

Proof. (1) If $\Delta = 0$, the assertion is contained in Theorem 33.

(2) If $\Delta \neq 0$, the inverse of the given mapping exists; the inverse mapping is also linear and is defined by formulas of form (4). Let $M(x_1, x_2, x_3)$ be any point of a line given in α by the equation

$$\rho x_1 + u_2x_2 + u_3x_3 = 0 \quad (7)$$

Assume further that $M'(x'_1, x'_2, x'_3)$ in α' is the corresponding point of M ; the coordinates of M and M' are related by equations (4). Multiplying both sides of (7) by ρ and substituting for $\rho x_1, \rho x_2, \rho x_3$ the expressions (4), we obtain

$$u'_1x'_1 + u'_2x'_2 + u'_3x'_3 = 0 \quad (8)$$

* Note that if $\text{Rank } C = 2$, then the points M' fill the entire line. But if $\text{Rank } C = 1$, all the points M' coincide and the image is only one point.

where

$$\begin{aligned}u'_1 &= c'_{11}u_1 + c'_{21}u_2 + c'_{31}u_3 \\u'_2 &= c'_{12}u_1 + c'_{22}u_2 + c'_{32}u_3 \\u'_3 &= c'_{13}u_1 + c'_{23}u_2 + c'_{33}u_3\end{aligned}$$

Thus, if M lies in α on a line given by equation (7), M' lies in α' on a line given by (8). ■

Combining Theorems 32 and 34 with the definition of projectivities of planes, we obtain the following

Theorem 35. *If the determinant of a linear mapping of a plane α into a plane α' does not vanish, then this mapping is a projectivity of α into α' .*

We shall show later that, conversely, every projectivity is a linear mapping. Before proving this important theorem, we shall obtain the following auxiliary

Theorem 36. *Let M_1, M_2, M_3, M_4 be any four points of a plane α no three of which are collinear and let M'_1, M'_2, M'_3, M'_4 be similarly any four points of a plane α' no three of which are collinear. Then there is a linear mapping of α into α' with a nonzero determinant that transforms M_1, M_2, M_3, M_4 into M'_1, M'_2, M'_3, M'_4 , respectively.*

Proof. Let x_{1k}, x_{2k}, x_{3k} be the coordinates of one of the points M_k ($k = 1, 2, 3, 4$) in a certain homogeneous coordinate system introduced in α and let $x'_{1k}, x'_{2k}, x'_{3k}$ be the coordinates of a point M'_k ($k = 1, 2, 3, 4$) in a certain homogeneous coordinate system introduced in α' . We have to show that it is possible to choose the parameters of the linear mapping

$$\left. \begin{aligned}\rho'x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho'x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho'x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3\end{aligned} \right\} \quad (1)$$

in such a way that its determinant is not zero and that it maps M_1, M_2, M_3, M_4 into M'_1, M'_2, M'_3, M'_4 , respectively. For this we must obviously show that from the equations

$$\left. \begin{aligned}\rho'_k x'_{1k} &= c_{11}x_{1k} + c_{12}x_{2k} + c_{13}x_{3k} \\ \rho'_k x'_{2k} &= c_{21}x_{1k} + c_{22}x_{2k} + c_{23}x_{3k} \\ \rho'_k x'_{3k} &= c_{31}x_{1k} + c_{32}x_{2k} + c_{33}x_{3k}\end{aligned} \right\} \quad (\alpha)$$

($k = 1, 2, 3, 4$) it is possible to find the parameters c_{ik} , with $\Delta \neq 0$, and the quantities ρ'_k (where ρ'_k is the factor appearing in the left hand side of (1) corresponding to the homogeneous coordinates of M_k and M'_k chosen).

Note first that none of the third-order determinants of the matrix

$$\begin{pmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \\ x_{14} & x_{24} & x_{34} \end{pmatrix} \quad (\beta)$$

is zero. Indeed, if, for example,

$$\begin{vmatrix} x_{11} & x_{21} & x_{31} \\ x_{12} & x_{22} & x_{32} \\ x_{13} & x_{23} & x_{33} \end{vmatrix} = 0,$$

then there exist linear relations

$$u_1 x_{11} + u_2 x_{21} + u_3 x_{31} = 0$$

$$u_1 x_{12} + u_2 x_{22} + u_3 x_{32} = 0$$

$$u_1 x_{13} + u_2 x_{23} + u_3 x_{33} = 0$$

and, accordingly, the points M_1 , M_2 , M_3 would lie on the line $u_1 x_1 + u_2 x_2 + u_3 x_3 = 0$, which contradicts the hypothesis of the theorem. It can be shown in exactly the same way that none of the third-order determinants of the matrix, similar to (β) , composed of the coordinates x'_{1k} , x'_{2k} , x'_{3k} is zero.

Let us now examine equations (α) . Setting $k = 1, 2, 3$, we write the three equations

$$\left. \begin{aligned} c_{11}x_{11} + c_{12}x_{21} + c_{13}x_{31} &= \rho'_1 x'_{11} \\ c_{11}x_{12} + c_{12}x_{22} + c_{13}x_{32} &= \rho'_2 x'_{12} \\ c_{11}x_{13} + c_{12}x_{23} + c_{13}x_{33} &= \rho'_3 x'_{13} \end{aligned} \right\} \quad (\gamma)$$

obtained from the first equation in (α) . This is a system of linear equations with unknowns c_{11} , c_{12} , c_{13} . In view of the preceding result, the determinant of system (γ) denoted by D is not zero. Hence

$$c_{11} = \frac{1}{D} \begin{vmatrix} \rho'_1 x'_{11} & x_{21} & x_{31} \\ \rho'_2 x'_{12} & x_{22} & x_{32} \\ \rho'_3 x'_{13} & x_{23} & x_{33} \end{vmatrix}, \quad c_{12} = \frac{1}{D} \begin{vmatrix} x_{11} & \rho'_1 x'_{11} & x_{31} \\ x_{12} & \rho'_2 x'_{12} & x_{32} \\ x_{13} & \rho'_3 x'_{13} & x_{33} \end{vmatrix}$$

$$c_{13} = \frac{1}{D} \begin{vmatrix} x_{11} & x_{21} & \rho'_1 x'_{11} \\ x_{12} & x_{22} & \rho'_2 x'_{12} \\ x_{13} & x_{23} & \rho'_3 x'_{13} \end{vmatrix}$$

or

$$\left. \begin{aligned} c_{11} &= \frac{\rho'_1 x'_{11} X_{11} + \rho'_2 x'_{12} X_{12} + \rho'_3 x'_{13} X_{13}}{D} \\ c_{12} &= \frac{\rho'_1 x'_{11} X_{21} + \rho'_2 x'_{12} X_{22} + \rho'_3 x'_{13} X_{23}}{D} \\ c_{13} &= \frac{\rho'_1 x'_{11} X_{31} + \rho'_2 x'_{12} X_{32} + \rho'_3 x'_{13} X_{33}}{D} \end{aligned} \right\} \quad (\delta_1)$$

where X_{ik} is the cofactors of x_{ik} in D . Similarly, using the second and third relations in (α) , we obtain

$$\left. \begin{aligned} c_{21} &= \frac{\rho'_1 x'_{21} X_{11} + \rho'_2 x'_{22} X_{12} + \rho'_3 x'_{23} X_{13}}{D} \\ c_{22} &= \frac{\rho'_1 x'_{21} X_{21} + \rho'_2 x'_{22} X_{22} + \rho'_3 x'_{23} X_{23}}{D} \\ c_{23} &= \frac{\rho'_1 x'_{21} X_{31} + \rho'_2 x'_{22} X_{32} + \rho'_3 x'_{23} X_{33}}{D} \end{aligned} \right\} \quad (\delta_2)$$

and

$$\left. \begin{aligned} c_{31} &= \frac{\rho'_1 x'_{31} X_{11} + \rho'_2 x'_{32} X_{12} + \rho'_3 x'_{33} X_{13}}{D} \\ c_{32} &= \frac{\rho'_1 x'_{31} X_{21} + \rho'_2 x'_{32} X_{22} + \rho'_3 x'_{33} X_{23}}{D} \\ c_{33} &= \frac{\rho'_1 x'_{31} X_{31} + \rho'_2 x'_{32} X_{32} + \rho'_3 x'_{33} X_{33}}{D} \end{aligned} \right\} \quad (\delta_3)$$

If we substitute the right hand sides of (δ_1) , (δ_2) , (δ_3) into equations (α) for $k = 4$, take $\rho'_4 = 1$ for simplicity and arrange terms, we have

$$\left. \begin{aligned} \frac{x'_{11} D_1}{D} \rho'_1 + \frac{x'_{12} D_2}{D} \rho'_2 + \frac{x'_{13} D_3}{D} \rho'_3 &= x'_{14}, \\ \frac{x'_{21} D_1}{D} \rho'_1 + \frac{x'_{22} D_2}{D} \rho'_2 + \frac{x'_{23} D_3}{D} \rho'_3 &= x'_{24}, \\ \frac{x'_{31} D_1}{D} \rho'_1 + \frac{x'_{32} D_2}{D} \rho'_2 + \frac{x'_{33} D_3}{D} \rho'_3 &= x'_{34} \end{aligned} \right\} \quad (\varepsilon)$$

where

$$\left. \begin{aligned} D_1 &= X_{11}x_{14} + X_{21}x_{24} + X_{31}x_{34} \\ D_2 &= X_{12}x_{14} + X_{22}x_{24} + X_{32}x_{34} \\ D_3 &= X_{13}x_{14} + X_{23}x_{24} + X_{33}x_{34} \end{aligned} \right\} \quad (\zeta)$$

The expressions D_1 , D_2 , D_3 show that these quantities are third-order determinants of matrix (β) . Accordingly, $D_1 \neq 0$, $D_2 \neq 0$, $D_3 \neq 0$. If we set

$$D' = \begin{vmatrix} x'_{11} & x'_{12} & x'_{13} \\ x'_{21} & x'_{22} & x'_{23} \\ x'_{31} & x'_{32} & x'_{33} \end{vmatrix}$$

then the determinant A of system (ε) can be written in the form

$$A = \frac{D'D_1D_2D_3}{D^3}$$

Whence $A \neq 0$ and hence the system (ε) is consistent. The solution of this system is

$$\begin{aligned} \rho'_1 &= \frac{D}{D'D_1} \begin{vmatrix} x'_{14} & x'_{12} & x'_{13} \\ x'_{24} & x'_{22} & x'_{23} \\ x'_{34} & x'_{32} & x'_{33} \end{vmatrix}, & \rho'_2 &= \frac{D}{D'D_2} \begin{vmatrix} x'_{11} & x'_{14} & x'_{13} \\ x'_{21} & x'_{24} & x'_{23} \\ x'_{31} & x'_{34} & x'_{33} \end{vmatrix} \\ \rho'_3 &= \frac{D}{D'D_3} \begin{vmatrix} x'_{11} & x'_{12} & x'_{14} \\ x'_{21} & x'_{22} & x'_{24} \\ x'_{31} & x'_{32} & x'_{34} \end{vmatrix} \end{aligned}$$

Since the determinants in terms of which $\rho'_1, \rho'_2, \rho'_3$ are expressed are not zero (this fact was demonstrated in the beginning of the proof), it follows that $\rho'_1 \neq 0, \rho'_2 \neq 0, \rho'_3 \neq 0$. Having found $\rho'_1, \rho'_2, \rho'_3$, the parameters c_{ik} can be uniquely determined from relations (δ) .

Linear mapping (1) with the obtained c_{ik} is the desired mapping because (i) it transforms points M_1, M_2, M_3, M_4 into points M'_1, M'_2, M'_3, M'_4 and (ii) its determinant is not zero; the latter result follows from Theorem 33. ■

In § 106 we proved Theorem 25, according to which a projectivity of one plane into another is uniquely determined by four pairs of corresponding points (subject to certain conditions regarding their location). The preceding results enable us to state a stronger theorem, namely

Theorem 37. *Given four points M_1, M_2, M_3, M_4 of a plane α , no three of which are collinear, and four points M'_1, M'_2, M'_3, M'_4 of a plane α' , also no three of which are collinear, there is a unique projectivity of α into α' that maps M_1, M_2, M_3, M_4 into M'_1, M'_2, M'_3, M'_4 , respectively.*

This theorem is a direct consequence of Theorems 35 and 36.

We shall prove, in passing, an analogous theorem concerning projectivities of lines.

Theorem 37a. *Given three distinct points M_1, M_2, M_3 of a line a and three distinct points M'_1, M'_2, M'_3 of a line a' , there exists only one projectivity of a into a' that maps M_1, M_2, M_3 into M'_1, M'_2, M'_3 , respectively.*

Proof. Since the uniqueness of such a mapping was proved in Theorem 15, we need only now show its existence.

Through line a draw a plane α and through a' a plane α' . Take in α any line u passing through M_3 (but not coinciding with a) and take any two points P, Q on this line; take in α' any line u' passing

through M'_3 (not coinciding with a') and any two points P' , Q' on this line. According to Theorem 37, there is a projectivity of α into α' that maps M_1, M_2, P, Q into M'_1, M'_2, P', Q' . By Theorem 24, this mapping transforms a projectively into a' in such a way that M_1 is mapped into M'_1 , M_2 into M'_2 , and M_3 into M'_3 ; the latter result follows from the fact that M_3 is common to both a and u , and M'_3 is common to the corresponding lines a' and u' . This proves the theorem. ■

We shall show later, in § 140, how to obtain a projectivity of a line into a line by means of three pairs of corresponding points without taking recourse to the mapping of planes.

We can now easily prove one of the most important theorems of projective geometry.

Theorem 38. *Every projectivity of a plane into a plane is a linear mapping with a nonvanishing determinant.*

Proof. The proof is very short. Let $M' = f(M)$ be any projectivity of a plane α into a plane α' . Choose in α four points M_1, M_2, M_3, M_4 no three of which are collinear. Let M'_1, M'_2, M'_3, M'_4 be the corresponding points in α' . In view of Theorem 23a, M'_1, M'_2, M'_3, M'_4 are also noncollinear. According to Theorem 36, there is a linear mapping $M' = L(M)$ of α into α' with a nonvanishing determinant and this mapping transforms M_1, M_2, M_3, M_4 into M'_1, M'_2, M'_3, M'_4 . Theorems 35 and 25 imply that the mappings $M' = f(M)$ and $M' = L(M)$ are not distinct. ■

Theorem 38 settles the question regarding analytic representation for projectivities, formulated at the beginning of the present section.

Every projectivity is represented in projective coordinates by the equations

$$\left. \begin{aligned} \rho' x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho' x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho' x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3 \end{aligned} \right\} \quad (1)$$

with a nonvanishing determinant

$$\Delta = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} \neq 0$$

§ 110. All the theorems obtained in the preceding subsection can be extended in a natural way to the three-dimensional case. We only formulate the main theorem.

Given a projectivity $M' = f(M)$ of a space Π into a space Π' , the projective coordinates x'_1, x'_2, x'_3, x'_4 of a point M' are expressed in terms of projective coordinates x_1, x_2, x_3, x_4 of point M by means of the linear

relations

$$\begin{aligned}\rho'x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 + c_{14}x_4 \\ \rho'x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 + c_{24}x_4 \\ \rho'x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3 + c_{34}x_4 \\ \rho'x'_4 &= c_{41}x_1 + c_{42}x_2 + c_{43}x_3 + c_{44}x_4\end{aligned}$$

with constant coefficients c_{ik} ; moreover,

$$\Delta = \begin{vmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \\ c_{31} & c_{32} & c_{33} & c_{34} \\ c_{41} & c_{42} & c_{43} & c_{44} \end{vmatrix} \neq 0$$

§ 111. We can now obtain an analytic representation for projectivities of a line into another line.

Let a and a' be two projective lines, and let $M' = f(M)$ be a projectivity of a into a' . We introduce a system of homogeneous projective coordinates on a by means of three points A_1, A_2, E , where A_1, A_2, E have, respectively, the coordinates $(0, 1)$, $(1, 0)$, $(1, 1)$ (see the end of § 100). Similarly, by means of three points A'_1, A'_2, E' homogeneous projective coordinates are introduced on a' . Our aim is to obtain equations that connect the coordinates of M and $M' = f(M)$.

We shall show that the coordinates (x_1, x_2) of M and the coordinates (x'_1, x'_2) of $M' = f(M)$ are connected by the equations

$$\left. \begin{aligned}\rho'x'_1 &= c_{11}x_1 + c_{12}x_2 \\ \rho'x'_2 &= c_{21}x_1 + c_{22}x_2\end{aligned} \right\} \quad (*)$$

where the coefficients $c_{11}, c_{12}, c_{21}, c_{22}$ defining the mapping are constants and $\rho' \neq 0$ is any number; the determinant

$$\Delta = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} \neq 0$$

To prove the result, draw a plane α through a and a plane α' through a' . Then introduce in α a homogeneous projective coordinate system such that A_1, A_2 and E have, respectively, the coordinates $(0, 1, 0)$, $(1, 0, 0)$ and $(1, 1, 0)$. Thus the line a is represented by the equation $x_3 = 0$, that is, is one of the lines of the coordinate trihedron; moreover, if M is any point of a and $(x_1, x_2, 0)$ are its coordinates in α , then x_1, x_2 coincide with the homogeneous coordinates of M in the coordinate system that was introduced on a (in order to define in α the coordinate system just mentioned, the two vertices of the coordinate trihedron should be taken at A_1 and A_2 and the third

vertex A_3 should be chosen arbitrarily and the unit point chosen somewhere on the line A_3E ; see § 101). Similarly, we introduce homogeneous projective coordinates on a' .

Having completed the above procedure, take any three points A, B, C on a and their corresponding point A', B', C' , under the mapping $M' = f(M)$, on a' . Further, take points D, G in α such that D, G, C lie on one line which is distinct from a ; similarly, choose points D', G' in α' such that D', G', C' lie on one line which is distinct from a' . By Theorem 37, there is a projectivity $M' = F(M)$ of α into α' that maps A, B, D, G respectively into A', B', D', G' . Evidently, the points of a are mapped into points of a' , and, in accordance with Theorem 24, the mapping $M' = F(M)$ which transforms line a into line a' is a projectivity. It is now easy to obtain formulas expressing coordinates of the point $M' = F(M)$ of a' in terms of the coordinates of the point M of a . To do this, we first write the known formulas

$$\left. \begin{aligned} \rho' x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho' x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho' x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3 \end{aligned} \right\} \quad (**)$$

expressing the coordinates of M' in terms of the coordinates of M for any location of M in α and then set $x_3 = 0$. Then for any x_1, x_2 the last equation in $(**)$ necessarily implies that $x'_3 = 0$ (because a is mapped into a'); accordingly, $c_{31} = c_{32} = 0$ and therefore, with $x_3 = 0, x'_3 = 0$, we have

$$\begin{aligned} \rho' x'_1 &= c_{11}x_1 + c_{12}x_2 \\ \rho' x'_2 &= c_{21}x_1 + c_{22}x_2 \end{aligned}$$

These are the desired equations connecting the coordinates of M and $M' = F(M)$ of lines a and a' , respectively, that is, the analytic representation of the mapping $M' = F(M)$. But it can be easily shown that the mapping $M' = F(M)$ of a into a' is not different from the given mapping $M' = f(M)$. Indeed, the mapping $M' = f(M)$ transforms a point C of a into point C' of a' . To establish this, consider the mapping $M' = F(M)$ of the entire plane α into plane α' ; this mapping transforms the lines AB and DG into lines $A'B'$ and $D'G'$ which means that it maps C , the point of intersection of AB and DG , into C' , the point of intersection of $A'B'$ and $D'G'$. Thus both $M' = F(M)$ and $M' = f(M)$ map the points A, B, C of a into points A', B', C' of a' . This, in view of Theorem 15, implies that the mappings $M' = F(M)$ and $M' = f(M)$ coincide. Thus, formulas $(*)$ give an analytic representation for any given projectivity $M' = f(M)$ of a line a into a line a' .

That the determinant

$$\Delta = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix}$$

does not vanish is a consequence of the fact that the projectivity in question is one-to-one. Indeed, if $\Delta = 0$, equations (*) do not determine x_1, x_2 for a given x'_1, x'_2 , that is, the mapping defined by formulas (*) is not one-to-one, contrary to the hypothesis. This completely proves the proposition.

Thus the problem of analytic representation for projectivities have been solved fully. The results obtained may be formulated in short thus: projectivities of projective manifolds are always represented in homogeneous projective coordinates by linear homogeneous equations.

§ 112. In many cases it turns out to be convenient to use analytic representations for projectivities in nonhomogeneous coordinates. These representations can be easily obtained from formulas expressing projectivities in homogeneous coordinates.

Recall that if x_1, x_2 are homogeneous projective coordinates of a point M of a line, then the quantity $x = x_1/x_2$ is the nonhomogeneous projective coordinate of the same point (see § 100). Let there be given a projectivity $M' = f(M)$ of a line a into a line a' . Then we know that the homogeneous coordinates of M and M' are related by the equations

$$\begin{aligned} \rho' x'_1 &= c_{11}x_1 + c_{12}x_2 \\ \rho' x'_2 &= c_{21}x_1 + c_{22}x_2 \end{aligned} \quad \left(\Delta = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} \neq 0 \right)$$

Dividing the first equation by the second and setting $x_1/x_2 = x$, $x'_1/x'_2 = x'$, we obtain

$$x' = \frac{c_{11}x + c_{12}}{c_{21}x + c_{22}} \quad (*)$$

which is the desired relation between the nonhomogeneous coordinates x and x' of M and M' . Introducing new coefficients $c_{11} = \alpha$, $c_{12} = \beta$, $c_{21} = \gamma$, $c_{22} = \delta$, we can write the last relation as

$$x' = \frac{\alpha x + \beta}{\gamma x + \delta} \quad (\Delta = \alpha\delta - \beta\gamma \neq 0) \quad (1)$$

Thus, if a projectivity of a line a into line a' is given, then the nonhomogeneous projective coordinates of the points of a' are expressed in terms of the nonhomogeneous projective coordinates of corresponding points of a by means of a linear-fractional function with a nonzero determinant.

Analytic representations in nonhomogeneous coordinates for projectivities of a plane into a plane and of space into space may be similarly obtained.

For instance, if $M' = f(M)$ is a projectivity of a plane α into a plane α' , then the homogeneous projective coordinates of M and M' satisfy equations (1) of § 109. Dividing the first and second equations by the third and setting $x_1/x_3 = x$, $x_2/x_3 = y$, $x'_1/x'_3 = x'$, $x'_2/x'_3 = y'$ and introducing new coefficients, we obtain the desired equations connecting the nonhomogeneous coordinates (x, y) and (x', y') of points M and M' in the form

$$\begin{aligned} x' &= \frac{a_1x + b_1y + c_1}{\alpha x + \beta y + \gamma} \\ y' &= \frac{a_2x + b_2y + c_2}{\alpha x + \beta y + \gamma} \end{aligned} \quad \left(\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ \alpha & \beta & \gamma \end{vmatrix} \neq 0 \right) \quad (2)$$

Similarly, if $M' = f(M)$ is a projectivity of space Π into space Π' , then the coordinates x', y', z' of the point M' are expressed in terms of the projective coordinates x, y, z of the point M by means of the equations

$$\begin{aligned} x' &= \frac{a_1x + b_1y + c_1z + d_1}{\alpha x + \beta y + \gamma z + \delta} \\ y' &= \frac{a_2x + b_2y + c_2z + d_2}{\alpha x + \beta y + \gamma z + \delta} \\ z' &= \frac{a_3x + b_3y + c_3z + d_3}{\alpha x + \beta y + \gamma z + \delta} \end{aligned} \quad \left(\Delta = \begin{vmatrix} a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \\ \alpha & \beta & \gamma & \delta \end{vmatrix} \neq 0 \right)$$

§ 113. We shall conclude the present section by considering an important particular case of the projectivity of one-dimensional manifolds.

Suppose that a projective one-dimensional manifold Π (a line, a linear pencil or a pencil of planes) is mapped projectively into itself in such a way that any element p is mapped into an element $p' = f(p)$. If the given mapping is performed twice in succession, the element p is first mapped into element $p' = f(p)$ and then into an element $p'' = f(p')$. Generally, p'' will not coincide with p .

Given an element p , if the element $p'' = f(p')$, where $p' = f(p)$, coincides with p , then the mapping $p' = f(p)$ is called an *involution*. The involution is a particular case of a projectivity which will now be examined.

The involutory character of a mapping $p' = f(p)$ may be expressed in one of the following ways:

- (1) $f(f(p)) = p$ for any p ;
- (2) $p = f(p')$, besides $p' = f(p)$, for any p , that is, the inverse mapping coincides with the given mapping.

Both these characterizations follow at once from the definition of an involution (namely, from the relation $p'' = p$).

Suppose that the one-dimensional manifold Π between whose elements an involution $p' = f(p)$ is given is either a pencil of lines

or a pencil of planes. Take any line a subject to the sole condition that if Π is a pencil of lines, then a lies in its plane and does not pass through the centre of the pencil, and if Π is a pencil of planes, then a does not intersect its axis. Let M denote the point of intersection of line a and element p of Π and let M' denote the point of intersection of a with element p' . The correspondence $M' = F(M)$, like $p' = f(p)$, between the points of a which assigns M' to M is evidently an involution. It is also obvious that the properties of involutions $p' = f(p)$ and $M' = F(M)$ are identical. Therefore, in examining involutions of one-dimensional manifolds, it is sufficient to deal only with the case where this manifold is a line.

Let $M' = F(M)$ be a projectivity of a line a into itself. Introduce on line a a system of (nonhomogeneous) projective coordinates. Then, as we know, the coordinates x and x' of the points M and $M' = F(M)$ satisfy the relation

$$x' = \frac{\alpha x + \beta}{\gamma x + \delta} \quad (\Delta = \alpha\delta - \beta\gamma \neq 0) \quad (*)$$

First, let us try to find the conditions which the coefficients α , β , γ , δ must satisfy in order for the mapping $M' = F(M)$ to be an involution. To do this, we solve (*) for x and obtain

$$x = \frac{-\delta x' + \beta}{\gamma x' - \alpha} \quad (**)$$

which is the analytic representation for the mapping that is the inverse of the given mapping. As noted above, an involution is characterized by the fact that this mapping coincides with its inverse. Comparing (*) and (**), we immediately see that the two mappings coincide either if $\delta = -\alpha$ or if $\delta = \alpha$, $\beta = 0$, $\gamma = 0$. When $\delta = \alpha$, $\beta = 0$, $\gamma = 0$, the mapping is an identity ($x' = x$). The identity involution will be dropped from our consideration. Then the condition $\delta = -\alpha$ is necessary and sufficient for the mapping (*) to be an involution.

Thus involutions are those projectivities which can be expressed in the form

$$x' = \frac{\alpha x + \beta}{\gamma x - \alpha} \quad (\Delta = -\alpha^2 - \beta\gamma \neq 0) \quad (***)$$

Now we shall establish a few simple theorems concerning involutions.

Theorem 39. *Let $M' = F(M)$ be a projectivity of a line a into itself. If there is a point M_0 which does not coincide with its image $M'_0 = F(M_0)$ but is mapped into itself by another application of the given mapping, then all the points do likewise and hence $M' = F(M)$ is an involution.*

Proof. Let $x' = (\alpha x + \beta)/(\gamma x + \delta)$ be the analytic representation for the mapping $M' = F(M)$ in any system of projective coordinates

on line a , and let x_0, x'_0 be the coordinates of M_0 and M'_0 . According to the hypothesis of the theorem,

$$x'_0 = \frac{\alpha x_0 + \beta}{\gamma x_0 + \delta} \quad \text{and} \quad x_0 = \frac{\alpha x'_0 + \beta}{\gamma x'_0 + \delta} \quad (x_0 \neq x'_0)$$

or

$$\gamma x_0 x'_0 + \delta x'_0 - \alpha x_0 - \beta = 0$$

$$\gamma x_0 x'_0 + \delta x_0 - \alpha x'_0 - \beta = 0$$

Subtracting the two equations, we find that

$$\delta(x'_0 - x_0) + \alpha(x'_0 - x_0) = 0$$

and, since $x'_0 - x_0 \neq 0$, it follows that $\delta = -\alpha$. ■

Theorem 40. *If an involution of the projective line into itself has fixed points, they are two in number.*

Proof. Let $x' = (\alpha x + \beta)/(\gamma x - \alpha)$ be the analytic representation for a certain involution. Obviously, the coordinates of fixed points (for which $x' = x$) are given by the equation

$$x = \frac{\alpha x + \beta}{\gamma x - \alpha}$$

or

$$\gamma x^2 - 2\alpha x - \beta = 0$$

If $\gamma \neq 0$, this equation is a quadratic equation with discriminant $\alpha^2 + \beta\gamma = -\Delta \neq 0$. Thus, this equation has either two complex roots, in which case the involution has no fixed points, or two distinct real roots, in which case there are two fixed points. But if $\gamma = 0$, then, setting $-\beta/\alpha = a$, we can write the relation between x and x' in the form $x' = -x + a$. In this case, the involution obviously has two fixed points, $x = a/2$ and $x = \infty$. ■

An involution that does not leave any point of the line fixed is called *elliptic*. Elliptic involutions are characterized by the condition $\Delta = -\alpha^2 - \beta\gamma > 0$.

An involution that leaves two points fixed is called *hyperbolic*. For hyperbolic involutions $\Delta = -\alpha^2 - \beta\gamma < 0$.

Sometimes the mapping $x' = (\alpha x + \beta)/(\gamma x - \alpha)$ is referred to as a parabolic involution if $\Delta = -\alpha^2 - \beta\gamma = 0$. But we know that a linear fractional mapping with a vanishing determinant is not a one-to-one mapping and therefore does not belong to the class of mapping under consideration.*

* That $x' = (\alpha x + \beta)/(\gamma x + \delta)$ with $\Delta = \alpha\delta - \beta\gamma = 0$ is not a one-to-one mapping may also be shown, besides the general consideration given above thus: if $\Delta = 0$, then $\alpha \div \gamma = \beta \div \delta = q$ and $x' = (\alpha x + \beta)/(\gamma x + \delta) = q(\gamma x + \delta)/(\gamma x + \delta) = q$. Accordingly, the same point with coordinate $x' = q$ is assigned to every point.

Theorem 41. *If $M' = f(M)$ is a hyperbolic involution with fixed points A and B , then the pair of points M, M' divides the pair A, B harmonically.*

Proof. We choose a system of nonhomogeneous projective coordinates on the line in such a way that A has the coordinate $x = 0$ and B has the (symbolic) coordinate $x = \infty$. Let

$$x' = \frac{\alpha x + \beta}{\gamma x - \alpha} \quad (*)$$

be the analytic representation of the involution $M' = f(M)$ in the coordinate system chosen. Since A is a fixed point with $x = 0$, we must have $x' = 0$ implying $\beta = 0$. Since B is also a fixed point, it follows that $x' \rightarrow \infty$ as $x \rightarrow \infty$, implying $\gamma = 0$. Thus $(*)$ assumes the form

$$x' = -x$$

Accordingly, $(x + x')/2 = 0$, that is, A is the projective centre of the segment xx' , and this implies that the pair M, M' divides the pair A, B harmonically (see § 97, Property (4) of homogeneous coordinates, and § 95, where the projective centre was defined). ■

This theorem can also be formulated in the following manner.

A hyperbolic involution $M' = f(M)$ with fixed points A, B is a harmonic mapping into each other of mutually complementary segments with common endpoints A and B (see § 93 and, in particular, the Remark at the end of that subsection).

Theorem 42. *An involution is uniquely determined by two distinct pairs of corresponding points.*

For the proof, it suffices to observe that the involution $x' = (\alpha x + \beta)/(\gamma x - \alpha)$ is determined by its numerical coefficients of which three are independent. But since a proportional change in the coefficients of a linear-fractional function does not affect this function, it follows that for determining an involution only two ratios $\alpha \div \beta \div \gamma$ need be known and for this two conditions suffice. Such conditions can be obtained by prescribing two pairs of corresponding points. If the coordinates of the given points are x_1, x'_1 and x_2, x'_2 , the ratios of the parameters of that involution under which these points are corresponding points may be found from

$$x_1^\bullet = \frac{\alpha x_1 + \beta}{\gamma x_1 - \alpha}, \quad x_2^\bullet = \frac{\alpha x_2 + \beta}{\gamma x_2 - \alpha}$$

or

$$\gamma x_1 x_1^\bullet - \alpha (x_1 + x_1^\bullet) - \beta = 0$$

$$\gamma x_2 x_2^\bullet - \alpha (x_2 + x_2^\bullet) - \beta = 0$$

Indeed, they yield

$$\alpha \div \beta \div \gamma = \begin{vmatrix} x_1 x'_1 & 1 \\ x_2 x'_2 & 1 \end{vmatrix} \div \begin{vmatrix} x_1 + x'_1 & x_1 x'_1 \\ x_2 + x'_2 & x_2 x'_2 \end{vmatrix} \div \begin{vmatrix} x_1 + x'_1 & 1 \\ x_2 + x'_2 & 1 \end{vmatrix}$$

The ratios $\alpha \div \beta \div \gamma$ remain undetermined only if all the determinants in the right-hand side of the last relation are zero. But then $x_1 + x'_1 = x_2 + x'_2$ and $x_1 x'_1 = x_2 x'_2$. Accordingly, either $x_1 = x_2$ and $x'_1 = x'_2$ or $x_1 = x'_2$ and $x_2 = x'_1$; both cases, however, are ruled out in view of the hypothesis of the theorem.

Thus, given any two distinct pairs of points M_1, M'_1 , and M_2, M'_2 , there is always a unique involution that maps M_1 into M'_1 and M_2 into M'_2 . ■

Observe that according to Theorem 15, an arbitrary projectivity is determined by three pairs of corresponding points, but, in contrast, an involution requires for its determination only two pairs. This is clearly due to the fact that the number of parameters in a general projectivity exceeds by one the number of parameters in an involution.

We end this section by one more remark.

The theorems just proved concern the involution of a line. If in the formulations of these theorems the words "point of a line" are everywhere replaced by the words "element of a one-dimensional manifold", then the theorems concerning the involutions of one-dimensional manifolds are obtained.

5.10. Transformation formulas for projective coordinates.

Cross ratio of four elements

§ 114. Suppose that two different systems of homogeneous projective coordinates are introduced on a projective line a ; one of the systems will be called the first system and the other the second. A point M of the line has coordinates (x_1, x_2) in the first system and (x'_1, x'_2) in the second. We pose the following problem: derive formulas expressing the coordinates (x'_1, x'_2) in terms of the coordinates (x_1, x_2) . The reader should note that in essence this problem was solved in the preceding section.

Indeed, the relationship between the coordinates of points that correspond under a projectivity of a line into another line was obtained in § 111. Let us examine the identity mapping of the line a into itself, that is, a mapping under which every point of a retains its position. The identity mapping is obviously a projectivity. Therefore the coordinates (x_1, x_2) of point M in the first system and the coordinates (x'_1, x'_2) of its image, that is, of the same point M in the second system, must satisfy exactly the same equations which were obtained

in § 111, namely

$$\begin{aligned} \rho'x'_1 &= c_{11}x_1 + c_{12}x_2 \\ \rho'x'_2 &= c_{21}x_1 + c_{22}x_2 \end{aligned} \quad \left(\Delta = \begin{vmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{vmatrix} \neq 0 \right) \quad (*)$$

These are the desired formulas. The numerical values of the c 's can be obtained for each concrete transformation by the conditions determining the transformation in question.

If the first equation of $(*)$ is divided by the second equation and the notations are changed thus: $c_{11} = \alpha$, $c_{12} = \beta$, $c_{21} = \gamma$, $c_{22} = \delta$, we obtain the formula for the transformation of nonhomogeneous projective coordinates (on a line):

$$x' = \frac{\alpha x + \beta}{\gamma x + \delta} \quad (\Delta = \alpha\delta - \beta\gamma \neq 0) \quad (**)$$

The same arguments enable us to conclude that the earlier formulas expressing the relationship between the projective coordinates of points which correspond to each other under a projectivity of a plane into a plane or of space into a space are also true for the transformation of projective coordinates in a plane or in space. Naturally, when these formulas define the transformation of (homogeneous) coordinates, the quantities x_i and x'_i ($i = 1, 2, 3$, for the plane and $i = 1, 2, 3, 4$ for the space) are different coordinates of one and the same point.

§ 115. By establishing the formulas for the transformation of projective coordinates we have solved a problem of prime importance. Only now can we settle the question regarding the invariance of the functions of coordinates of points or the invariance of the relationships between the coordinates of points which one encounters in projective geometry. In particular, now we can define the concept of the *cross ratio of four elements* of a one-dimensional manifold, which proves to be of immense importance for projective geometry. We start with the cross ratio of four points of a line.

Let M_1, M_2, M_3, M_4 be four points of a projective line a on which we introduce a system of projective coordinates (for computational purposes nonhomogeneous coordinates are more convenient), and denote by t_1, t_2, t_3, t_4 the coordinates of points in question. We shall show that the quantity

$$\frac{t_3 - t_1}{t_2 - t_3} \div \frac{t_4 - t_1}{t_2 - t_4}$$

is independent of the choice of the coordinate system. Consider a new system of projective coordinates, apart from the one which was introduced at the beginning. If the coordinate of any point of a is t in the old system and t' in the new, then

$$t' = \frac{\alpha t + \beta}{\gamma t + \delta} \quad (\Delta = \alpha\delta - \beta\gamma \neq 0)$$

where $\alpha, \beta, \gamma, \delta$ are constants depending on the chosen coordinate systems. In particular, for the new coordinates t'_1, t'_2, t'_3 of the points M_1, M_2, M_3 , we have

$$t'_1 = \frac{\alpha t_1 + \beta}{\gamma t_1 + \delta}, \quad t'_2 = \frac{\alpha t_2 + \beta}{\gamma t_2 + \delta}, \quad t'_3 = \frac{\alpha t_3 + \beta}{\gamma t_3 + \delta}$$

Therefore

$$t'_3 - t'_1 = \frac{(\alpha\delta - \beta\gamma)(t_3 - t_1)}{(\gamma t_1 + \delta)(\gamma t_3 + \delta)}$$

$$t'_2 - t'_3 = \frac{(\alpha\delta - \beta\gamma)(t_2 - t_3)}{(\gamma t_2 + \delta)(\gamma t_3 + \delta)}$$

and

$$\frac{t'_3 - t'_1}{t'_2 - t'_3} = \frac{\gamma t_2 + \delta}{\gamma t_1 + \delta} \frac{t_3 - t_1}{t_2 - t_3}$$

Similarly,

$$\frac{t'_4 - t'_1}{t'_2 - t'_4} = \frac{\gamma t_2 + \delta}{\gamma t_1 + \delta} \frac{t_4 - t_1}{t_2 - t_4}$$

and hence

$$\frac{t'_3 - t'_1}{t'_2 - t'_3} \div \frac{t'_4 - t'_1}{t'_2 - t'_4} = \frac{t_3 - t_1}{t_2 - t_3} \div \frac{t_4 - t_1}{t_2 - t_4}$$

as required.

The quantity

$$(M_1 M_2 M_3 M_4) = \frac{t_3 - t_1}{t_2 - t_3} \div \frac{t_4 - t_1}{t_2 - t_4} \quad (*)$$

is known as the cross ratio of four points M_1, M_2, M_3, M_4 .

The preceding result shows that a cross ratio is determined only by the location of points M_1, M_2, M_3, M_4 ; a coordinate system is required only for calculating its actual numerical value and thus plays only a secondary role.

The following properties of the cross ratio are an immediate consequence of formula (*):

$$(1) \quad (M_1 M_2 M_3 M_4) = (M_3 M_4 M_1 M_2)$$

that is, the first and second pairs of points of a cross ratio may be interchanged without affecting the value of the cross ratio.

$$(2) \quad (M_1 M_2 M_4 M_3) = \frac{1}{(M_1 M_2 M_3 M_4)}$$

that is, if the points of a pair are interchanged, the value of the cross ratio is replaced by its reciprocal.

§ 116. We shall obtain some more theorems concerning cross ratio of four points.

Theorem 43. *Under a projectivity of a line a into line a' , the cross ratio of any set of points M_1, M_2, M_3, M_4 of a is the same as the cross ratio of the corresponding points M'_1, M'_2, M'_3, M'_4 of a' .*

To prove this result, we introduce coordinate systems on lines a and a' . If t_i are the coordinates of M_i and t'_i those of M'_i , then

$$(M_1M_2M_3M_4) = \frac{t_3 - t_1}{t_2 - t_3} \div \frac{t_4 - t_1}{t_2 - t_4}$$

$$(M'_1M'_2M'_3M'_4) = \frac{t'_3 - t'_1}{t'_2 - t'_3} \div \frac{t'_4 - t'_1}{t'_2 - t'_4}$$

According to § 112, the coordinates t and t' of points M and M' which correspond under the given projectivity satisfy the relation $t' = (\alpha t + \beta)/(\gamma t + \delta)$. With the aid of this relation the equality $(M'_1M'_2M'_3M'_4) = (M_1M_2M_3M_4)$ may be established in exactly the same way as in the preceding article. ■

A particular case of the preceding theorem is

Theorem 44. *If a and a' are two lines in a plane α and S is any point in α lying neither on a nor on a' , then the cross ratio of any four points M_1, M_2, M_3, M_4 of a equals the cross ratio of their projections M'_1, M'_2, M'_3, M'_4 from the centre S to a' .*

That this theorem is indeed a particular case of the preceding theorem follows from the fact that a central projection is a particular case of a projectivity (see § 103). Theorem 44 may also be formulated thus:

Any four rays m_1, m_2, m_3, m_4 of a plane pencil determine, by intersecting any line in the plane of the given pencil, four points the value of whose cross ratio is the same for all the lines.

This value is called the *cross ratio of the rays* m_1, m_2, m_3, m_4 and is denoted by $(m_1m_2m_3m_4)$.

The cross ratio of four elements of a pencil of planes may be similarly defined: if $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ are four planes passing through one line, then by intersecting any line in space these planes determine four points the value of whose cross ratio is the same for all the lines. This value is known as the *cross ratio of the four planes* $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ and is denoted by $(\alpha_1\alpha_2\alpha_3\alpha_4)$.

The following theorem plays the same role for the projectivities of pencils (of rays or planes) as Theorem 43 does for the projectivities of lines.

Theorem 45. *Under a projectivity of a pencil into another pencil the cross ratio of any set of four elements of one pencil equals the cross ratio of the corresponding elements belonging to the other pencil.*

Proof. Let x be any element of one pencil and $x' = f(x)$ the corresponding element of the other pencil. Let the elements of the first pencil be intersected by a line a and those of the second by line a' , and let M be the point of intersection of a and x and M' that

of a' and x' . Consider a mapping of a into a' which assigns $M' = F(M)$ to M . Both the mappings $M' = F(M)$ and $x' = f(x)$ are projectivities. Therefore, if M_1, M_2, M_3, M_4 are points of intersection of a with arbitrary elements x_1, x_2, x_3, x_4 of the first pencil and M'_1, M'_2, M'_3, M'_4 are points of intersection of a' with the corresponding elements x'_1, x'_2, x'_3, x'_4 belonging to the second pencil,

$$(M'_1 M'_2 M'_3 M'_4) = (M_1 M_2 M_3 M_4)$$

(see Theorem 43). Furthermore, according to the definition of the cross ratio of four elements of a pencil, we have

$$(x_1 x_2 x_3 x_4) = (M_1 M_2 M_3 M_4)$$

$$(x'_1 x'_2 x'_3 x'_4) = (M'_1 M'_2 M'_3 M'_4)$$

Accordingly,

$$(x'_1 x'_2 x'_3 x'_4) = (x_1 x_2 x_3 x_4)$$

as required. ■

The following theorem, which contains both Theorems 43 and 45, is as easy to prove.

Theorem 46. *Under a projectivity of a one-dimensional manifold into another one-dimensional manifold the cross ratio of any set of four elements of the first manifold equals the cross ratio of the corresponding elements of the second manifold.*

This theorem contains as special cases the mappings of a line into a pencil of rays, of a line into a pencil of planes and of a pencil of rays into a pencil of planes.

Theorems 46 and 24 yield

Theorem 47. *If a plane α is projectively mapped into a plane α' , the cross ratio of any set of four elements belonging to a one-dimensional manifold of α equals the cross ratio of the four corresponding elements of α' .*

Theorems 46 and 28 yield a similar theorem regarding the projectivities of a space into another space.

The above results may be formulated in short as follows:

A cross ratio is invariant under projectivities.

§ 117. Take any four points A, B, C, D on a line a and another four points A', B', C', D' on the same line or on another line a' . What are the conditions under which there exists a projectivity of a into a' that maps the points A, B, C, D into the points A', B', C', D' ?

Theorem 43 states that a necessary condition for this is the equality of cross ratios of the sets A, B, C, D and A', B', C', D' . It is easy to find that this condition is also sufficient.

Indeed, supposing that $(ABCD) = (A'B'C'D')$, let $M' = f(M)$ be the projectivity of a into a' that maps the points A, B, C , into the points A', B', C' . The existence of such a projectivity is ensured

by Theorem 37a. Let $f(D) = D^*$. According to Theorem 43, $(ABCD) = (A'B'C'D^*)$, and therefore $(A'B'C'D^*) = (A'B'C'D')$. This result in conjunction with the definition of a cross ratio immediately implies that the point D^* coincides with the point D' . Thus under the condition $(ABCD) = (A'B'C'D')$ there does exist a projectivity of line a into line a' , namely the mapping $M' = f(M)$ that transforms A, B, C, D into A', B', C', D' .

In accordance with the definition given earlier, a system of points $M_1, M_2, M_3, \dots, M_n$ of a line a is projectively equivalent to the system of points $M'_1, M'_2, M'_3, \dots, M'_n$ of a line a' if there is a projectivity of a into a' which maps $M_1, M_2, M_3, \dots, M_n$ into $M'_1, M'_2, M'_3, \dots, M'_n$, respectively.

Arguing in the same way, the following proposition may be easily proved:

In order that a system of points $M_1, M_2, \dots, M_n$ of a line a be projectively equivalent to the system of points $M'_1, M'_2, \dots, M'_n$ of another line a' , it is necessary and sufficient that the cross ratio of any set of four points M_p, M_q, M_r, M_s of the first system be equal to the cross ratio of the corresponding points M'_p, M'_q, M'_r, M'_s of the second system.

The cross ratio which enables us to characterize the projective equivalence is the basic invariant of projective geometry just like the distance between points characterizing congruence is the basic invariant of elementary geometry.

§ 118. If on any line the pair of points A, B divides the pair of points C, D harmonically, then $(ABCD) = -1$.

To prove this result, first note that the cross ratio of four distinct points can never be $+1$. In fact, if

$$(M_1M_2M_3M_4) = \frac{x_3 - x_1}{x_2 - x_3} \div \frac{x_4 - x_1}{x_2 - x_4} = 1$$

we have $(x_3 - x_1)/(x_2 - x_3) = (x_4 - x_1)/(x_2 - x_4)$, which necessarily implies that $x_3 = x_4$ if $x_1 \neq x_2$, and this contradicts the assumption that the points are distinct.

Further, according to the definition of a harmonic set of points, if the pair A, B divides the pair C, D harmonically, there is a complete four-point $PQRS$ located with respect to the points A, B, C, D as shown in Fig. 116. From centre S project A, B, C, D to the line PQ . Since a cross ratio is invariant under a projection (see Theorem 44), we find that $(ABCD) = (PQED)$. Now again project A, B, C, D to PQ but this time from the centre R . Then $(ABCD) = (QPED)$. But according to § 115,

$$(QPED) = \frac{1}{(PQED)}$$

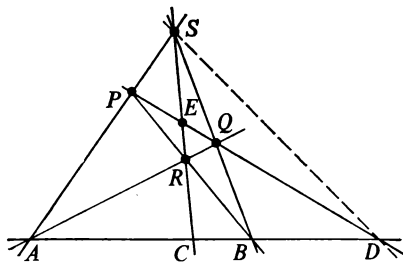


Figure 116

Thus $(ABCD) = (PQED)$ and $(ABCD) = \frac{1}{(PQED)}$, which yields

$$(ABCD)^2 = 1$$

and $(ABCD) = \pm 1$. But since a cross ratio cannot be equal to $+1$, it follows that

$$(ABCD) = -1$$

as required. If this result is combined with the definition of the cross ratio of four elements of a pencil, the following general theorem can be immediately obtained.

Theorem 48. *If the pair of elements x, y of a one-dimensional manifold divides another pair z, t of it harmonically, then $(xyzt) = -1$.*

§ 119. Now we shall obtain some new formulas expressing the cross ratio in projective coordinates.

Let a system of nonhomogeneous projective coordinates be given with the origin at the point O and the points at infinity ∞_x and ∞_y of the axes (see § 98). We consider any four collinear points M_1, M_2, M_3, M_4 in the plane and pose the following problem: Express the cross ratio $(M_1M_2M_3M_4)$ in terms of the coordinates (x^1, y^1) , (x^2, y^2) , (x^3, y^3) , (x^4, y^4) of the points in question.

With this in mind, project M_1, M_2, M_3, M_4 from the point ∞_y to the x -axis and denote by M'_1, M'_2, M'_3, M'_4 the respective projections. They have coordinates $(x^1, 0)$, $(x^2, 0)$, $(x^3, 0)$, $(x^4, 0)$ and their cross ratio can be written in the form

$$(M'_1M'_2M'_3M'_4) = \frac{x^3 - x^1}{x^2 - x^3} \div \frac{x^4 - x^1}{x^2 - x^4}$$

If $M''_1M''_2M''_3M''_4$ denote the projections of points M_1, M_2, M_3, M_4 from ∞_x to the y -axis, we similarly have

$$(M''_1M''_2M''_3M''_4) = \frac{y^3 - y^1}{y^2 - y^3} \div \frac{y^4 - y^1}{y^2 - y^4}$$

By Theorem 44,

$$(M_1 M_2 M_3 M_4) = (M'_1 M'_2 M'_3 M'_4) \\ (M_1 M_2 M_3 M_4) = (M''_1 M''_2 M''_3 M''_4)$$

Accordingly, any one of the following formulas

$$\left. \begin{aligned} (M_1 M_2 M_3 M_4) &= \frac{x^3 - x^1}{x^2 - x^3} \div \frac{x^4 - x^1}{x^2 - x^4} \\ (M_1 M_2 M_3 M_4) &= \frac{y^3 - y^1}{y^2 - y^3} \div \frac{y^4 - y^1}{y^2 - y^4} \end{aligned} \right\} \quad (A)$$

is applicable for the quantity $(M_1 M_2 M_3 M_4)$.

In order to obtain the cross ratio $(M_1 M_2 M_3 M_4)$ in homogeneous coordinates, one has only to replace the nonhomogeneous coordinates x^i, y^i of each point M_i ($i = 1, 2, 3, 4$) in the above formulas by the ratios of homogeneous coordinates $x_1^i/x_3^i, x_2^i/x_3^i$ of this point; proceeding in this way, we obtain

$$\left. \begin{aligned} (M_1 M_2 M_3 M_4) &= \frac{\begin{vmatrix} x_1^3 & x_1^1 \\ x_3^3 & x_3^1 \end{vmatrix}}{\begin{vmatrix} x_1^2 & x_1^3 \\ x_3^2 & x_3^3 \end{vmatrix}} \div \frac{\begin{vmatrix} x_1^4 & x_1^1 \\ x_3^4 & x_3^1 \end{vmatrix}}{\begin{vmatrix} x_1^2 & x_1^3 \\ x_3^2 & x_3^3 \end{vmatrix}} \\ (M_1 M_2 M_3 M_4) &= \frac{\begin{vmatrix} x_2^3 & x_2^1 \\ x_3^3 & x_3^1 \end{vmatrix}}{\begin{vmatrix} x_2^2 & x_2^3 \\ x_3^2 & x_3^3 \end{vmatrix}} \div \frac{\begin{vmatrix} x_2^4 & x_2^1 \\ x_3^4 & x_3^1 \end{vmatrix}}{\begin{vmatrix} x_2^2 & x_2^3 \\ x_3^2 & x_3^3 \end{vmatrix}} \end{aligned} \right\} \quad (B)$$

(in these formulas superscripts indicate the numbers of the points and the subscripts the numbers of the coordinates).

The cross ratio of four points of a line in space is expressed in projective coordinates by exactly the same formulas; only one more formula is added.

Let us note one special way of writing the cross ratio that follows from formulas (B).

Let P and Q be two distinct points in the plane whose homogeneous coordinates are p_i and q_i ($i = 1, 2, 3$). It may easily be shown that any point T with homogeneous coordinates

$$t_i = p_i + tq_i \quad (i = 1, 2, 3)$$

lies on the line PQ . In fact, if

$$u_1 x_1 + u_2 x_2 + u_3 x_3 = \sum u_i x_i = 0$$

is the equation of the line PQ , then $\sum u_i p_i = 0$ and $\sum u_i q_i = 0$, but then also

$$\sum u_i t_i = \sum u_i (p_i + tq_i) = \sum u_i p_i + t \sum u_i q_i = 0$$

that is, the coordinates of T satisfy the equation of the line PQ .

We arbitrarily assign two values $t = \lambda$ and $t = \mu$ to the parameter t and express according to formula (B) the cross ratio of four points P, Q, L, M with homogeneous coordinates $p_i, q_i, p_i + \lambda q_i, p_i + \mu q_i$; using, for example, the first formula in (B), we have

$$(PQLM) = \frac{\begin{vmatrix} p_1 + \lambda q_1 & p_1 \\ p_3 + \lambda q_3 & p_3 \end{vmatrix}}{\begin{vmatrix} q_1 & p_1 + \lambda q_1 \\ q_3 & p_3 + \lambda q_3 \end{vmatrix}} : \frac{\begin{vmatrix} p_1 + \mu q_1 & p_1 \\ p_3 + \mu q_3 & p_3 \end{vmatrix}}{\begin{vmatrix} q_1 & p_1 + \mu q_1 \\ q_3 & p_3 + \mu q_3 \end{vmatrix}} =$$

$$= \frac{\lambda \begin{vmatrix} q_1 & p_1 \\ q_3 & p_3 \end{vmatrix}}{\begin{vmatrix} q_1 & p_1 \\ q_3 & p_3 \end{vmatrix}} : \frac{\mu \begin{vmatrix} q_1 & p_1 \\ q_3 & p_3 \end{vmatrix}}{\begin{vmatrix} q_1 & p_1 \\ q_3 & p_3 \end{vmatrix}} = \frac{\lambda}{\mu}$$

It should be pointed out here that these calculations yield a definite number only if $\begin{vmatrix} q_1 & p_1 \\ q_3 & p_3 \end{vmatrix} \neq 0$. If, however, $\begin{vmatrix} q_1 & p_1 \\ q_3 & p_3 \end{vmatrix} = 0$, then the second formula in (B) should be used. The computations yield a definite number, which coincides with the previous result only if $\begin{vmatrix} q_2 & p_2 \\ q_3 & p_3 \end{vmatrix} \neq 0$. Both $\begin{vmatrix} q_1 & p_1 \\ q_3 & p_3 \end{vmatrix}$ and $\begin{vmatrix} q_2 & p_2 \\ q_3 & p_3 \end{vmatrix}$ cannot be zero because otherwise $p_1 \div p_2 \div p_3 = q_1 \div q_2 \div q_3$, which is impossible since the points P, Q are distinct. Thus

$$(PQLM) = \lambda/\mu \quad (C)$$

Such a way of writing a cross ratio is very simple and convenient in applications; it is evident that this form is also valid if the coordinates of points P, Q, L, M are taken in space.

We conclude this section by deriving a formula for the cross ratio of four rays of a pencil.

Consider a pencil with the centre $S(x_0, y_0)$. The equation of any ray of the pencil can be expressed in a nonhomogeneous coordinate system as

$$y - y_0 = k(x - x_0) \quad (*)$$

Let m_1, m_2, m_3, m_4 be four rays of the given pencil and k_1, k_2, k_3, k_4 the corresponding values of the parameter k in equation (*), and $M_1(x_1, 0), M_2(x_2, 0), M_3(x_3, 0), M_4(x_4, 0)$ the points of intersection of the rays m_1, m_2, m_3, m_4 with the x -axis. By the definition of the cross ratio of four points,

$$(M_1M_2M_3M_4) = \frac{x_3 - x_1}{x_2 - x_3} \div \frac{x_4 - x_1}{x_2 - x_4}$$

and by the definition of the cross ratio of four rays of a pencil, $(m_1m_2m_3m_4) = (M_1M_2M_3M_4)$. Accordingly,

$$(m_1m_2m_3m_4) = \frac{x_3 - x_1}{x_2 - x_3} \div \frac{x_4 - x_1}{x_2 - x_4}$$

For $y = 0$, equation (*) implies

$$x_i = x_0 - y_0/k_i \quad (i = 1, 2, 3, 4)$$

Substituting these values of x_1, x_2, x_3, x_4 in the preceding expression for $(m_1 m_2 m_3 m_4)$ and after some computations, we have the formula

$$(m_1 m_2 m_3 m_4) = \frac{k_3 - k_1}{k_2 - k_3} \div \frac{k_4 - k_1}{k_2 - k_4} \quad (\text{D})$$

expressing the cross ratio of four rays in terms of their parameters.

From formula (D) one may easily obtain the analytic representation for the projectivity of a pencil into another pencil, that is, obtain the form of the function $k' = f(k)$, where k and k' are parameters determining the projectively corresponding rays of two pencils.

In fact, if k_1, k_2, k_3 are parameters of three rays of the first pencil and k'_1, k'_2, k'_3 are parameters of the corresponding rays of the second pencil, then

$$\frac{k'_3 - k'_1}{k'_2 - k'_3} \div \frac{k' - k'_1}{k'_2 - k'} = \frac{k_3 - k_1}{k_2 - k_3} \div \frac{k - k_1}{k_2 - k}$$

because the cross ratio is, by Theorem 45, invariant under a projectivity of pencils. Solving for k' , we obtain

$$k' = \frac{\alpha k + \beta}{\gamma k + \delta} \quad (\text{E})$$

where $\alpha, \beta, \gamma, \delta$ are constants (depending on $k_1, k_2, k_3, k'_1, k'_2, k'_3$).

Thus a projectivity of a pencil into a pencil is represented analytically by a linear-fractional function. The determinant $\Delta = \alpha\delta - \beta\gamma$ of this function is not zero because otherwise the mapping defined by (E) would not be one-to-one.

5.11. The principle of duality

§ 120. First we shall consider the principle of duality on the projective plane. In the present section we shall formulate and prove one of the most striking features of projective geometry, known as the *principle of duality*. At first we shall confine our discussion to the two-dimensional case and present the main features of the duality principle for projective geometry on the plane.

In two-dimensional projective geometry, we consider objects of two kinds, points and planes. The incidence relationships between these objects are set out in the projective axioms of Groups I, II, III; of these axioms only the Axioms I.1-I.3, I.9, II.1-II.6 and III are axioms concerning two-dimensional geometry, and the remaining axioms (that is, Axioms I.4-I.8) concern the space. But, as was made clear during the previous discussion, in order to obtain

theorems for two-dimensional projective geometry without employing the three-dimensional constructions, one has to adjoin Desargues' theorem to Axioms I.1-I.3, I.9, II.4-II.6 and III. The axioms just mentioned and Desargues' theorem will together be referred to as the *two-dimensional projective axioms*. An analysis of the foundations of projective geometry reveals that a proposition can be associated with every two-dimensional projective axiom such that one of the two associated propositions, if suitably formulated, becomes the second proposition by replacing the word "point" with the word "line" and "line" with "point". This feature, which in essence is the duality principle on the plane (exact formulation will be given later), will become clear to the reader after we have actually performed the mentioned association.

[We start with the axioms of the first group. These axioms define the mutual incidence relationships between points and lines which are expressed in the usual language: "a point lies on a line" or "a line passes through a point". Instead of using these expressions, it is more convenient for us to use the expressions "a point belongs to a line" or "a line belongs to a point". Using these expressions, we rewrite Axioms I.1-I.3 and I.9 on the left side and the proposition which is associated in the above sense with the given axiom on the right side, opposite each axiom. Two such propositions will henceforth be called *dual*.

(I.1) Given two points A and B , there is a line a belonging to both A and B .

(I.2) Given two distinct points A and B , there is at most one line belonging to both A and B .

(I.3) At least three points belong to every line. There are at least three points that do not belong to a given line.

(I.9) Any two lines have a common point.

Desargues' theorem. Let three points A , B , C do not belong to one line and, similarly, let three points A' , B' , C' not be-

Given two lines a and b , there is a point A belonging to both a and b . (This is nothing but Axiom I.9.)

Given two distinct lines a and b , there is at most one point belonging to both a and b . (This proposition follows immediately from Axiom I.2.)

At least three lines belong to every point. There are at least three lines not belonging to one point. (This assertion can be proved by means of Axioms I.1-I.3.)

Any two points have a common line. (This is a restatement of Axiom I.1.)

Let three lines a , b , c not belong to one point and, similarly, three lines a' , b' , c' not belong to one point. Suppose

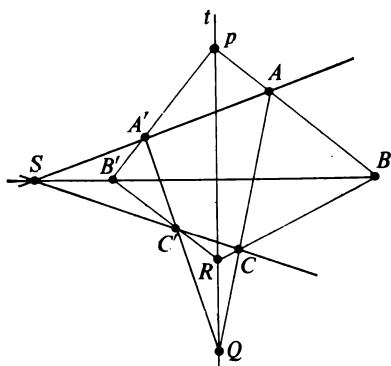


Figure 117

long to one line. Suppose further that the line belonging to A and B has a common point P with the line belonging to A' and B' ; the line belonging to B , C has a common point Q with the line belonging to B' , C' , and the line belonging to A , C has a common point R with the line belonging to A' , C' . Then, if P , Q , R belong to one line t , the three lines, one of which belongs to A , A' , the second to B , B' and the third to C , C' , have a common point S (Fig. 117).

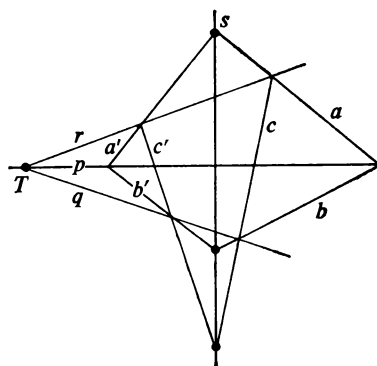


Figure 118

further that the point belonging to a , b has a common line p with the point belonging to a' , b' ; the point belonging to b , c has a common line q with the point belonging to b' , c' , and the point belonging to a , c has a common line r with the point belonging to a' , c' . Then, if p , q , r belong to one point T , the three points, one of which belongs to the lines a , a' , the second to b , b' and the third to c , c' have a common line s (Fig. 118). (This proposition is nothing but Theorem 2, the converse of Desargues' theorem.)

Thus with every two-dimensional projective axiom of the first group it is indeed possible to associate a proper (that is, following from these axioms) assertion so that the associated propositions are dual.

Let us now examine the axioms of the second group, Axioms II.1-II.6.

The basic concept used in Axioms II.1-II.6 is that of dividing pairs of points of a line; the definition of dividing pairs of lines passing through one point (see § 89) is given in terms of this concept. The axioms of the second group with their dual propositions are given below; the validity of the dual propositions follow at once from the axioms of Groups I and II and the definition of dividing pairs of lines.

(II.1) Given three distinct points A, B, C belonging to one line u , there is a point D belonging to u such that the pair A, B divides the pair C, D .

If the pair A, B divides the pair C, D , then all four points A, B, C, D are distinct.

(II.2) If the pair A, B divides the pair C, D , then the pair B, A divides the pair C, D and the pair C, D divides the pair A, B .

(II.3) Given four distinct points A, B, C, D belonging to one line u , two dividing pairs can always be formed from these points in only one way.

(II.4) Let A, B, C, D, E be points belonging to a line u . If the pairs C, D and C, E divide the pair A, B , then the pair D, E does not divide the pair A, B .

(II.5) Let A, B, C, D, E be points belonging to a line u . If the pairs C, D and C, E do not divide the pair A, B , then the pair D, E does not divide the pair A, B either.

(II.6) Let a, b and c, d be two pairs of lines belonging to one point S , and u, u' two lines not belonging to S . Suppose further that A, B, C, D are points belonging to u and, respectively, to a, b, c, d , and A', B', C', D' are points belonging to u' and, respectively, to a, b, c, d . Then, if the pair A, B divides the pair C, D , the pair A', B' divides the pair C', D' .

Given three distinct lines a, b, c belonging to one point U , there is a line d belonging to U such that the pair a, b divides the pair c, d .

If the pair a, b divides the pair c, d , then all four lines a, b, c, d are distinct.

If the pair a, b divides the pair c, d , then the pair b, a divides the pair c, d and the pair c, d divides the pair a, b .

Given four distinct lines a, b, c, d belonging to one point U , two dividing pairs can always be formed from these lines in only one way.

Let a, b, c, d, e be lines belonging to a point U . If the pairs c, d and c, e divide the pair a, b , then the pair d, e does not divide the pair a, b .

Let a, b, c, d, e be lines belonging to a point U . If the pairs c, d and c, e do not divide the pair a, b , then the pair d, e does not divide the pair a, b either.

Let A, B and C, D be two pairs of points belonging to one line s , and U, U' two points not belonging to s . Suppose further that a, b, c, d are lines belonging to U and, respectively, to A, B, C, D and a', b', c', d' are lines belonging to U' and, respectively, to A, B, C, D . Then, if the pair a, b divides the pair c, d , the pair a', b' divides the pair c', d' .

Thus dual propositions are indeed associated with the axioms of the second group.

Consider, finally, the axiom of continuity, Axiom III.

In order to be able to formulate Axiom III (Dedekind's axiom), we had to define the linear orderings of points on the deleted projective line. Let us recall this definition.

Taking any line a , we choose a point U on this line and for the remaining points of this line we introduce a "betweenness" relationship by assuming that of the three points A, B, C the point C lies between A and B if the pair A, B is divided by the pair C, U . We say that for the set of points of line a which is obtained by omitting U a linear ordering has been defined if this set is ordered subject to the following condition: as soon as the point C lies between the points A and B in the sense of the established ordering, the point C also lies between A and B in the sense of the definition just accepted.

With the aim of formulating a proposition dual to Axiom III, we shall now define a linear ordering for the set of all lines, except one, that pass through one point.

Taking any point A , we choose a line u among all those lines which pass through A , and for the remaining lines we introduce a "betweenness" relationship by assuming that of the three lines a, b, c the line c passes between a and b if the pair a, b is divided by the pair c, u . We shall say that a linear ordering has been defined for the set of all lines, excluding the line u , that pass through A if this set is ordered subject to the following condition: as soon as c lies between the lines a and b in the sense of the established ordering, the line c lies between a and b also in the sense of the definition just given.

Now we can state Axiom III and its dual proposition.

(III) Let a be any line, U any point belonging to a , and let a linear ordering be introduced on the set of the remaining points belonging to a . If this set is classified into two classes such that (1) every point belongs to one and only one class, (2) every class contains points, and (3) any point of the first class precedes every point of the second, then either the first class contains a point which follows (in the sense of the ordering established) all the points of this class or the second class contains a point which precedes all its points.

Let A be any point, u any line belonging to A , and let a linear ordering be introduced on the set of remaining lines belonging to A . If this set is classified into two classes such that (1) every line belongs to one and only one class, (2) every class contains lines, and (3) any line of the first class precedes every line of the second, then either the first contains a line which follows (in the sense of the ordering established) all the lines of this class or the second contains a line which precedes all its lines.

The validity of the proposition dual to Axiom III may be easily established by means of the operation of section. In fact, let S be any point and u any line not passing through S . We associate with every line passing through S that point of u which belongs to this line. If in the set of all lines, excluding one line, passing through S as well as in the set of points corresponding to these lines a linear ordering is defined, then the corresponding elements of these two sets are either always in the same order or always in the reverse order. Therefore Dedekind's principle holds for the set of lines passing through S as soon as this principle holds for the set of points of line u , that is, the validity of the proposition dual to Axiom III follows from this axiom.

Thus, every axiom of two-dimensional projective geometry has indeed a dual proposition. On the basis of the above analysis the following principle may be enunciated.

Let there be given two sets of objects, called, respectively, points and lines, such that between these objects incidence and order relationships are established subject to the requirements of all the axioms of two-dimensional projective geometry. If the roles of these objects are interchanged, that is, if the objects of the first set are called lines while those of the second points and the mutual relationships between them are left unchanged, then all the requirements of the projective axioms will again be fulfilled.

§ 121. Obviously, we can develop projective geometry starting by either the axioms originally accepted or by their duals. From the logical point of view, in both cases we shall be working on the same problem.

If such a dual construction of projective geometry be carried out, then together with every projective theorem we shall obtain its dual, and then all the theorems can be paired in suitable formulation such that one proposition of the pair will lead to another proposition of the same pair simply by replacing the word "point" with the word "line", and vice versa. It is not difficult to point out an abstract formulation of theorems of projective geometry in which the mutually dual propositions will be combined into one. To do so, one has to refrain completely from using the words "point" and "line" replacing them with the words "object of the first kind" and "object of the second kind". Every theorem formulated in an abstract manner may then be interpreted in two ways, once by taking the objects of the first kind to mean points and the objects of the second kind to mean lines, and a second time by taking the objects of the first and second kind to mean lines and points, respectively. The dual theorems obtained by two interpretations, if assigned to some definite realization of projective geometry, will express, in general, different facts. For example, the assertion "two

objects of the first kind always define one and only one common object of the second kind" leads to two dual propositions: (1) that one and only one line always passes through two distinct points and (2) that two distinct lines always intersect at one point.

If the words "point" and "line" in these two propositions are understood in the same sense, then these propositions have different concrete meanings.

There are theorems, among them important ones, in projective geometry which were discovered at various periods of time and even in various epochs but which, being dual, coincide in the abstract logical construction of projective geometry. An example is provided by the famous theorems of Pascal and Brianchon (see § 143), which were discovered one hundred years apart but which coincide logically.

From the modern point of view the principle of duality does not appear as something surprising. It is easily detected if geometry is examined from the abstractly logical point of view. But at the beginning of the nineteenth century the discovery of the duality principle was a highly original and progressive feat; in particular, the principle of duality has played a leading role in the development of the abstract understanding of geometrical objects.

In our previous discussion the dual character of projective geometry constantly appeared in that the propositions regarding a system of points of a line were associated with similar propositions regarding elements of a pencil. Observe that there is no duality in elementary geometry. For example, the Euclidean points and lines are not dual with respect to their incidence relationships; indeed, in the Euclidean plane, where two points always have a common line, two lines do not always have a common point (the lines may be parallel). The duality is absent in order relationships also; namely, all points of a Euclidean line are situated in a linear order but the order of rays in a pencil is cyclic. One may also note the important differences in congruence relationships of segments and angles (which are completely absent in projective geometry); for instance, in the Euclidean plane triangles with corresponding congruent sides are equal, but triangles with corresponding congruent angles are, in general, not equal.

§ 122. It is natural that the duality in two-dimensional projective geometry has a certain analytic expression.

In order to obtain suitable association of the corresponding analytic relations, besides the dual association of statements of projective geometry, we use the coordinates of lines, which are defined below.

Suppose that a system of homogeneous projective coordinates is introduced in the plane. Then every point of the plane is determined by the ratio of three numbers x_1 , x_2 , x_3 and every line by

an equation of the form

$$u_1x_1 + u_2x_2 + u_3x_3 = 0 \quad (*)$$

We shall call the coefficients u_1, u_2, u_3 in equation (*) the coordinates of the line represented by this equation. Evidently, the coordinates u_1, u_2, u_3 are homogeneous because the three numbers u_1, u_2, u_3 and the three numbers $\rho u_1, \rho u_2, \rho u_3$ determine one and the same line. In other words, in order to determine a line it is enough to know the ratio $u_1 \div u_2 \div u_3$. It is also obvious that any three numbers u_1, u_2, u_3 are coordinates of some line except when all three numbers are zero.

From the preceding discussion it follows that if (x_1, x_2, x_3) are the coordinates of a point P and (u_1, u_2, u_3) are the coordinates of a line p , then the relation

$$u_1x_1 + u_2x_2 + u_3x_3 = 0$$

is the condition for both P and p to belong to each other. Thus we have the following dual propositions:

With fixed (u_1, u_2, u_3) and moving (x_1, x_2, x_3) , the equation

$$(*) \quad u_1x_1 + u_2x_2 + u_3x_3 = 0$$

determines all possible points belonging to the line (u_1, u_2, u_3) ; in this sense this is called the equation of the line (u_1, u_2, u_3) .

With fixed (x_1, x_2, x_3) and moving (u_1, u_2, u_3) , the equation

$$(*) \quad x_1u_1 + x_2u_2 + x_3u_3 = 0$$

determines all possible lines belonging to the point (x_1, x_2, x_3) ; in this sense it is called the equation of the point (x_1, x_2, x_3) .*

Next, note that if (p_1, p_2, p_3) and (q_1, q_2, q_3) are the coordinates of two points P and Q , then for any λ the numbers $p_1 + \lambda q_1, p_2 + \lambda q_2, p_3 + \lambda q_3$ are the coordinates of some point L of the line PQ . Indeed, let $u_1x_1 + u_2x_2 + u_3x_3 = 0$ be the equation of the line PQ . Since the coordinates of P and Q must satisfy this equation, we have $u_1p_1 + u_2p_2 + u_3p_3 = 0$ and $u_1q_1 + u_2q_2 + u_3q_3 = 0$. But then

$$u_1(p_1 + \lambda q_1) + u_2(p_2 + \lambda q_2) + u_3(p_3 + \lambda q_3) = 0$$

that is, the coordinates of L satisfy the equation of the line PQ , and hence L does lie on the line PQ .

Similarly, if (v_1, v_2, v_3) and (w_1, w_2, w_3) are the coordinates of two lines v and w , then for any λ the numbers $v_1 + \lambda w_1, v_2 + \lambda w_2, v_3 + \lambda w_3$ are the coordinates of some line l that passes through the point of intersection of v and w .

* By the way, for fixed x_1, x_2, x_3 and moving u_1, u_2, u_3 equation (*) is more frequently called the *equation of the pencil* with the centre (x_1, x_2, x_3) .

Indeed, let O be the point of intersection of the lines v and w and $x_1u_1 + x_2u_2 + x_3u_3 = 0$ be its equation; the coordinates of v and w must satisfy this equation, and, accordingly, $x_1v_1 + x_2v_2 + x_3v_3 = 0$ and $x_1w_1 + x_2w_2 + x_3w_3 = 0$. But then

$$x_1(v_1 + \lambda w_1) + x_2(v_2 + \lambda w_2) + x_3(v_3 + \lambda w_3) = 0$$

that is, the coordinates of line l satisfy the equation of point O and, therefore, l does pass through O .

It was shown in § 119 that the cross ratio of the points P, Q, L, M with the coordinates $p_i, q_i, p_i + \lambda q_i, p_i + \mu q_i$ ($i = 1, 2, 3$) is given by

$$(PQLM) = \lambda/\mu \quad (1)$$

In view of the principle of duality, the cross ratio of the lines v, w, l, m with coordinates $v_i, w_i, v_i + \lambda w_i, v_i + \mu w_i$ ($i = 1, 2, 3$) can be expressed by a similar formula

$$(vwl m) = \lambda/\mu \quad (2)$$

Formulas (1), (2) and Theorem 48 imply the following dual propositions:

If the points P, Q, L, M have respectively the coordinates $p_i, q_i, p_i + \lambda q_i, p_i + \mu q_i$ ($i = 1, 2, 3$), then the necessary and sufficient condition for the pairs P, Q and L, M to divide harmonically is

$$\lambda/\mu = -1$$

If the lines v, w, l, m have respectively the coordinates $v_i, w_i, v_i + \lambda w_i, v_i + \mu w_i$ ($i = 1, 2, 3$), then a necessary and sufficient condition for the pair v, w and l, m to divide harmonically is

$$\lambda/\mu = -1$$

Like in the examples cited, it is clear that in all other cases analytic relations corresponding to dual facts of projective geometry go over to each other by replacing the coordinates of points by those of lines, and vice versa.

§ 123. We now turn to the principle of duality in the projective space. In three-dimensional projective geometry we deal with objects of three kinds, called points, lines and planes, and two kinds of their mutual relationships: the incidence and betweenness relationships.

Instead of the expressions "a point lies in a plane" or "a plane passes through a point", used in visual geometry, we shall say "a point belongs to a plane" or "a plane belongs to a point"; instead of saying "a point lies on a line" or "a line passes through a point" we shall say "a point belongs to a line" or "a line belongs to a point", and instead of "a line lies in a plane" or "a plane passes through a line" we shall say "a line belongs to a plane" or "a plane belongs to a line".

If the Axioms I.1-I.9, which state the properties of mutual incidence of objects, are suitably formulated, then with each of these axioms a valid proposition (following from Axioms I.1-I.9) may be associated such that the two associated propositions reduce to each other when the word "point" is replaced by the word "plane" and "plane" by "point" (the word "line" remains unchanged). The propositions thus associated will be called *dual*.

We write Axioms I.1-I.9 together with their dual propositions, the proofs of which are left to the reader.

(I.1) Given two points A and B , there is a line a belonging to both A and B .

(I.2) Given two distinct points A and B , there is at most one line belonging to A and B .

(I.3) At least three points belong to every line. There are at least three points not belonging to a given line.

(I.4) Given three points A , B , C not belonging to one line, there is a plane α belonging to the points A , B , C . At least one point belongs to every plane.

(I.5) Given three points A , B , C not belonging to one line, there is at most one common plane belonging to them.

(I.6) If two points A , B belonging to a line a also belong to a plane α , then every point belonging to a also belongs to α .

(I.7) If a common point A belongs to two planes α , β , then at least one more common point B belongs to these planes.

(I.8) There are at least four points not belonging to a given plane.

(I.9) Any two lines belonging to one plane belong to a common point.

Given two planes α and β , there is a line a belonging to both α and β .

Given two distinct planes α and β , there is at most one line belonging to α and β .

At least three planes belong to every line. There are at least three planes not belonging to a given line.

Given three planes α , β , γ , not belonging to one line, there is a point A belonging to the planes α , β , γ . At least one plane belongs to every point.

Given three planes α , β , γ , not belonging to one line, at most one common point belongs to them.

If two planes α , β belonging to a line a also belong to the point A , then every plane belonging to a also belongs to A .

If a common plane α belongs to two points A , B , then at least one more common plane β belongs to these points.

There are at least four planes not belonging to a given point.

Any two lines belonging to one point belong to a common plane.

There is no need of describing in as much detail the propositions which are dual to the axioms of Group II and Axiom III. The man-

ner of their formulation is clear enough from the preceding exposition and their proofs require totally trivial arguments.

In view of the fact that all the propositions which are dual to Axioms I, II, III are valid (that is, are implied by these axioms themselves), we have the following principle of duality in space:

Let there be given three sets of objects, called points, lines and planes, between which incidence and order relationships are established such that the requirements of all the axioms of projective geometry are satisfied. If the roles of these objects are interchanged by calling the objects of the first set planes and those of the third set points (the objects of the second set remain unchanged) and if the mutual relationships between these objects are not changed, then all the requirements of the projective axioms will again be satisfied.

Evidently, three-dimensional projective geometry can be developed, like two-dimensional geometry, by starting with either the axioms accepted originally or their duals.

If such a dual construction of projective geometry is carried out, then together with every projective theorem we shall obtain its dual.

Of course, having proved a certain projective theorem, if we wish to obtain its dual, there is no need of actually proving this dual theorem; the formulation of the dual theorem can be obtained from the given theorem by replacing the words according to the following scheme:

point $\rightarrow$ plane

line $\rightarrow$ line

plane $\rightarrow$ point

Then its validity is established by the duality principle.

For definite geometric objects, the dual propositions express, in general, different concrete facts. For example, all the theorems regarding the figures composed of points and lines of one plane have as their duals the theorems regarding bodies composed of lines and planes passing through one point; in other words, geometry in the plane is dual to geometry of the bundle.

§ 124. Guided by the duality principle, we introduce the coordinates of the planes, apart from the coordinates of points. Namely, by coordinates of any plane α we mean the coefficients u_1, u_2, u_3, u_4 in the equation

$$u_1x_1 + u_2x_2 + u_3x_3 + u_4x_4 = 0$$

representing α in a certain system of homogeneous projective coordinates.

The coordinates (u_1, u_2, u_3, u_4) are obviously homogeneous because the four numbers u_1, u_2, u_3, u_4 and the four numbers $\rho u_1,$

$\rho u_2, \rho u_3, \rho u_4$ determine one and the same plane. Thus, to determine a plane it is enough to know the ratio $u_1 \div u_2 \div u_3 \div u_4$. Evidently, any four numbers u_1, u_2, u_3, u_4 are the coordinates of some plane except when all these four numbers are zero.

According to the preceding result, if (x_1, x_2, x_3, x_4) are the coordinates of a point P and (u_1, u_2, u_3, u_4) are the coordinates of a plane α , then

$$u_1x_1 + u_2x_2 + u_3x_3 + u_4x_4 = 0$$

is the condition for the point P and the plane α to belong to each other. Thus we have the following dual propositions:

For fixed (u_1, u_2, u_3, u_4) and current (x_1, x_2, x_3, x_4) the equation

$$u_1x_1 + u_2x_2 + u_3x_3 + u_4x_4 = 0$$

determines all possible points belonging to the plane (u_1, u_2, u_3, u_4) ; in this sense it is called the equation of the plane.

For fixed (x_1, x_2, x_3, x_4) and current (u_1, u_2, u_3, u_4) the equation

$$x_1u_1 + x_2u_2 + x_3u_3 + x_4u_4 = 0$$

determines all possible planes belonging to the point (x_1, x_2, x_3, x_4) ; in this sense it is called the equation of a bundle of planes.

Furthermore, the following assertions can be easily shown to be true:

If x_i, y_i, z_i ($i = 1, 2, 3, 4$) are the coordinates of three points X, Y, Z , then for any α, β, γ (except when $\alpha = \beta = \gamma = 0$) the equations

$$p_i = \alpha x_i + \beta y_i + \gamma z_i$$

determine the coordinates of a point P belonging to the plane XYZ ; therefore these equations are called the parametric equations of this plane.

If x_i, y_i ($i = 1, 2, 3, 4$) are the coordinates of two points X, Y , then for any α, β (except when $\alpha = \beta = 0$) the equations

$$(*) \quad p_i = \alpha x_i + \beta y_i$$

If $\alpha_i, \beta_i, \gamma_i$ ($i = 1, 2, 3, 4$) are the coordinates of three planes α, β, γ , then for any u, v, w (except when $u = v = w = 0$), the equations

$$\pi_i = u\alpha_i + v\beta_i + w\gamma_i$$

determine the coordinates of a plane π belonging to the common point of α, β, γ ; therefore these equations are called the parametric equations of a bundle of planes with centre at this point.

If α_i, β_i ($i = 1, 2, 3, 4$) are the coordinates of two planes α, β , then for any u, v (except when $u = v = 0$) the equations

$$(*) \quad \pi_i = u\alpha_i + v\beta_i$$

determine the coordinates of a point P belonging to the line XY ; therefore these equations are called the parametric equations of this line.

If we divide the equations (*) by α and set $\beta/\alpha = \lambda$, then the current coordinates p_i will be represented in terms of one parameter:

$$p_i = x_i + \lambda y_i$$

For different values of the parameter the equations in the first case define a set of points belonging to the line and in the second case a set of planes belonging to the line.

According to § 119, the cross ratio of four points X, Y, L, M with the coordinates $x_i, y_i, x_i + \lambda y_i, x_i + \mu y_i$ ($i = 1, 2, 3, 4$) is given by

$$(XYLM) = \lambda/\mu$$

This implies, in view of the duality principle, that the cross ratio of four planes $\alpha, \beta, \tau, \sigma$ with coordinates $\alpha_i, \beta_i, \alpha_i + t\beta_i, \alpha_i + s\beta_i$ ($i = 1, 2, 3, 4$) is given by

$$(\alpha\beta\tau\sigma) = t/s$$

Like the examples just cited, in other cases also the analytic expressions corresponding to dual facts reduce to each other if the coordinates of points are replaced by those of planes, and vice versa.

5.12. Algebraic curves and pencils.

Algebraic surfaces and bundles.

Complex projective plane and complex projective space

§ 125. One of the major subjects of investigation in plane projective geometry are the algebraic curves and their duals, according to the principle of duality, the algebraic pencils. These concepts are defined below.

An *algebraic curve* is the set of points whose homogeneous projective coordinates satisfy a homogeneous algebraic equation:

$$\sum a_{\alpha_1\alpha_2\dots\alpha_n} x_{\alpha_1} x_{\alpha_2} \dots x_{\alpha_n} = 0$$

$$(\alpha_1, \alpha_2, \dots, \alpha_n = 1, 2, 3)$$

An *algebraic pencil* is the set of lines whose homogeneous projective coordinates satisfy a homogeneous algebraic equation:

$$\sum a_{\alpha_1\alpha_2\dots\alpha_n} u_{\alpha_1} u_{\alpha_2} \dots u_{\alpha_n} = 0$$

$$(\alpha_1, \alpha_2, \dots, \alpha_n = 1, 2, 3)$$

where on the left side is written a homogeneous form of degree n ; the coefficients $a_{\alpha_1\alpha_2\dots\alpha_n}$ are assumed to be independent of the order in which the indices appear. The degree n of the equation is the *order* of the algebraic curve.

where on the left side is written a homogeneous form of degree n ; the coefficients $a_{\alpha_1\alpha_2\dots\alpha_n}$ are assumed to be independent of the order in which the indices appear. The degree n of the equation is the *class* of the algebraic pencil.

For $n = 1, 2, 3, \dots$, we respectively have:

a curve of the 1st order defined by the equation

$$a_1x_1 + a_2x_2 + a_3x_3 = 0$$

that is, the set of points belonging to one line; a curve of the 2nd order defined by the equation

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 + a_{33}x_3^2 = 0;$$

a curve of the 3rd order defined by the equation

$$\begin{aligned} & a_{111}x_1^3 + 3a_{112}x_1^2x_2 + 3a_{122}x_1x_2^2 \\ & + a_{222}x_2^3 + 3a_{223}x_2^2x_3 + 3a_{233}x_2x_3^2 \\ & + 3a_{113}x_1^2x_3 + 3a_{133}x_1x_3^2 \\ & + 6a_{123}x_1x_2x_3 + a_{333}x_3^3 = 0, \text{ etc.} \end{aligned}$$

a pencil of the 1st class defined by the equation

$$a_1u_1 + a_2u_2 + a_3u_3 = 0$$

that is, the set of lines belonging to one point; a pencil of the 2nd class defined by the equation

$$a_{11}u_1^2 + 2a_{12}u_1u_2 + a_{22}u_2^2 + 2a_{13}u_1u_3 + 2a_{23}u_2u_3 + a_{33}u_3^2 = 0;$$

a pencil of the 3rd class defined by the equation

$$\begin{aligned} & a_{111}u_1^3 + 3a_{112}u_1^2u_2 + 3a_{122}u_1u_2^2 \\ & + a_{222}u_2^3 + 3a_{223}u_2^2u_3 + 3a_{233}u_2u_3^2 \\ & + 3a_{113}u_1^2u_3 + 3a_{133}u_1u_3^2 \\ & + 6a_{123}u_1u_2u_3 + a_{333}u_3^3 = 0, \text{ etc.} \end{aligned}$$

According to the definition just given, algebraic curves are distinguished among curves in general by the form of their equations. A natural question could be: Is the form of an algebraic equation disturbed if one goes from one system of projective coordinates to another? In such a case it would be meaningless to introduce the concept of an algebraic curve in geometry. However, as can be easily shown, the form and degree of an algebraic equation are invariant under any transformation of projective coordinates. In fact, it is known that when the projective system of homogeneous coordinates is changed, the coordinates of points in the plane become homogeneous linear functions of the new coordinates and, in turn, the new coordinates are homogeneous linear functions of the old. But then it is obvious that in new coordinates also we obtain a homogeneous form and, moreover, of the same degree n as the original

form. Thus the concept of an algebraic curve and its order has a geometric meaning independent of the chosen coordinate system.

In order to demonstrate a similar property for the algebraic pencils, we must first obtain formulas according to which the homogeneous coordinates of lines correspond with a change in the projective coordinate system. To this end consider the equation

$$u_1x_1 + u_2x_2 + u_3x_3 = 0 \quad (*)$$

of any line and the equation

$$u'_1x'_1 + u'_2x'_2 + u'_3x'_3 = 0 \quad (**)$$

of the same line in new coordinates. Let

$$\left. \begin{aligned} \rho'x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho'x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho'x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3 \end{aligned} \right\} \quad (\alpha)$$

be the formulas connecting the new and old coordinates of points.

Since (*) and (**) determine the same line, we must have

$$u'_1x'_1 + u'_2x'_2 + u'_3x'_3 = \mu(u_1x_1 + u_2x_2 + u_3x_3)$$

for any x_1, x_2, x_3 . Replacing x'_1, x'_2, x'_3 in the last relation by their expressions in terms of x_1, x_2, x_3 , we obtain the identity

$$\frac{1}{\rho'} [(c_{11}u'_1 + c_{21}u'_2 + c_{31}u'_3)x_1 + (c_{12}u'_1 + c_{22}u'_2 + c_{32}u'_3)x_2 + (c_{13}u'_1 + c_{23}u'_2 + c_{33}u'_3)x_3] = \mu(u_1x_1 + u_2x_2 + u_3x_3).$$

Setting $\rho'\mu = \sigma$, the last identity yields

$$\left. \begin{aligned} \sigma u_1 &= c_{11}u'_1 + c_{21}u'_2 + c_{31}u'_3 \\ \sigma u_2 &= c_{12}u'_1 + c_{22}u'_2 + c_{32}u'_3 \\ \sigma u_3 &= c_{13}u'_1 + c_{23}u'_2 + c_{33}u'_3 \end{aligned} \right\} \quad (\beta)$$

which are the desired formulas connecting the old and new coordinates of lines. Thus the formulas for the transformation of coordinates of lines have the same structure as the formulas for the transformation of coordinates of points (and the determinant of transformation (β) is the same as that of transformation (α) and is hence $\neq 0$). Therefore the concepts of an algebraic pencil and its class, like that of an algebraic curve and its order, have a geometrical meaning independent of the chosen coordinate system.

It is clear that when the old variables in an algebraic equation are replaced by linear forms of the new variables, then an equation is obtained in which the coefficients, generally speaking, differ from the coefficients of the original equation. It is also clear that all the geometric properties of curves and pencils and all geometric quanti-

ties connected with them must be represented analytically by relationships between the coefficients of equations and by functions of these coefficients that do not change with a change in the projective coordinate system.

Thus, the problem of investigating algebraic curves and pencils in plane projective geometry is equivalent to the algebraic problem of investigating the invariants of homogeneous forms with three arguments.

One more remark is in order. An analysis of formulas (α) and (β) shows that the coordinates appearing in these formulas may also be assumed to belong to one system. Then the numbers x_1, x_2, x_3 and x'_1, x'_2, x'_3 in formula (α), for example, will not be different coordinates of one point but coordinates of different points M and M' . We know that a mapping of the projective plane into itself under which a point M goes to point M' determined by (α) is a projectivity. With this interpretation of formulas (α) and (β) (as formulas for projectivities) the invariance of the structure of equations of algebraic images with respect to transformations (α) and (β) means that under a projectivity algebraic images of any order or class are mapped into algebraic images of the same order or class.

Furthermore, it is obvious that if we perform a projectivity given by formulas (α) and simultaneously by these formulas define a change of projective coordinates, then an arbitrary algebraic image A in the system (x_1, x_2, x_3) and its projectively corresponding image A' in the system (x'_1, x'_2, x'_3) will have the same equations. Because algebraic images that can be projectively mapped into each other have identical equations in suitable coordinates, they have identical geometric properties too. This is in line with the feature of entire projective geometry where the figures which are mapped into each other by a projectivity are treated as equivalent (just as in elementary geometry those figures are treated as the same which coincide under motions).

§ 126. Any line contains at most n points of a curve of the n th order or else is entirely composed of points of this curve. To see this, let

$$\sum a_{\alpha_1 \alpha_2 \dots \alpha_n} x_{\alpha_1} x_{\alpha_2} \dots x_{\alpha_n} = 0$$

be the equation of some n th order curve, p_i, q_i ($i = 1, 2, 3$) be the coordinates of two points P and Q . The coordinates x_i ($i = 1, 2, 3$) of any point of the line PQ may be expressed as functions of parameter λ :

$$x_i = p_i + \lambda q_i \quad (i = 1, 2, 3)$$

These formulas determine common points of PQ and the given curve if λ satisfies the equation

$$\sum a_{\alpha_1 \alpha_2 \dots \alpha_n} (p_{\alpha_1} + \lambda q_{\alpha_1}) (p_{\alpha_2} + \lambda q_{\alpha_2}) \dots (p_{\alpha_n} + \lambda q_{\alpha_n}) = 0 \quad (*)$$

Suppose that Q is chosen subject to the condition

$$\sum a_{\alpha_1 \alpha_2 \dots \alpha_n} q_{\alpha_1} q_{\alpha_2} \dots q_{\alpha_n} \neq 0$$

(which is possible if the line is not entirely composed of the points of the given curve). In that case the left hand side of equation (*) contains λ^n and this equation is of degree n . Since to every real root λ_i there corresponds a point of intersection of the line PQ and the given algebraic curve and since the number of the real roots of (*) is at most n , it follows that the maximum number of common points of the line and the n th order curve is indeed n .

Analogously, an arbitrary point contains at most n lines of a pencil of class n or else all the lines belonging to the point belong to the pencil. In fact, let

$$\sum a_{\alpha_1 \alpha_2 \dots \alpha_n} u_{\alpha_1} u_{\alpha_2} \dots u_{\alpha_n} = 0$$

be the equation of a pencil of class n and let S be any point determined by the intersection of two lines v and w with the coordinates v_i and w_i ($i = 1, 2, 3$). The coordinates u_i ($i = 1, 2, 3$) of any line belonging to S may be expressed as a function of the parameter λ :

$$u_i = v_i + \lambda w_i \quad (i = 1, 2, 3)$$

These formulas determine the lines belonging to S and the given pencil provided λ satisfies the equation

$$\sum a_{\alpha_1 \alpha_2 \dots \alpha_n} (v_{\alpha_1} + \lambda w_{\alpha_1}) (v_{\alpha_2} + \lambda w_{\alpha_2}) \dots (v_{\alpha_n} + \lambda w_{\alpha_n}) = 0 \quad (**)$$

Let the line w be chosen subject to the condition

$$\sum a_{\alpha_1 \alpha_2 \dots \alpha_n} w_{\alpha_1} w_{\alpha_2} \dots w_{\alpha_n} \neq 0$$

(which is possible provided not all the lines belonging to S belong to the given pencil). In that case the left hand side of (**) contains λ^n and this equation is of degree n . Since to every real root λ_i there corresponds a line belonging to S and the given pencil and since the number of real roots of (**) is at most n , the maximum number of lines in the pencil passing through S is indeed n .

The propositions just proved might lead us to think that the order of an algebraic curve may be interpreted from the point of view of visual geometry as the maximum number of points of this curve belonging to one line and the class of a pencil as the maximum number of its lines belonging to one point. That such an interpretation is misleading can easily be demonstrated. Namely, there are curves of order n which have less than n common points with any given line. As an example, it is enough to cite the 2nd order curve $x_1^2 + x_2^2 + x_3^2 = 0$, which has no points whatsoever.

At the same time, the geometric interpretation for the order of a curve and the class of a pencil just mentioned will be valid if the

set of elements of the projective plane is augmented by adjoining new "imaginary" elements. The imaginary elements in geometry serve the same useful purpose as the imaginary numbers do in algebra because they enable us to simplify the formulation of many theorems. The principle of introducing imaginary elements on the projective plane is enunciated below.

§ 127. Let there be given two sets of objects, called, respectively, points and lines and let the incidence and order relationships between the elements of these two sets be defined satisfying the two-dimensional projective axioms (in other words, let the projective plane be given). Then, as we know, with all the points of the plane one can associate, according to some definite rule, in a one-to-one manner the ratios of real numbers x_1, x_2, x_3 , called the homogeneous projective coordinates of points, and with all the lines the ratios of real numbers u_1, u_2, u_3 , called projective coordinates of lines. By an *imaginary point* we shall mean any system of three complex numbers x_1, x_2, x_3 if at least one of these numbers is not zero and if the ratio of at least two of them is not a real number (the points (x_1, x_2, x_3) and $(\rho x_1, \rho x_2, \rho x_3)$, where ρ is any nonzero complex number, will be assumed to coincide). Under the same conditions, the triple u_1, u_2, u_3 of complex numbers will be called an *imaginary line*. Thus any triple of numbers can be treated both as a point and a line.

If the system of projective coordinates is changed, then the coordinates of all points change according to (α) , § 125, and those of lines according to (β) . We shall assume that the same formulas determine the coordinates of imaginary points and lines in any new projective coordinate system.

Thus the concepts of imaginary points and lines have an invariant meaning. Namely, we may state that imaginary points and lines are certain objects which are determined in a given projective coordinate system by triples of complex numbers with complex ratios; two triples of numbers in one coordinate system determine one and the same object provided the numbers constituting the triples are proportional; two triples of numbers in distinct coordinate systems determine one and the same object if they are connected by formulas (α) or (β) depending on whether the object in question is a point or a line.

For the augmented set of objects the incidence relationships are defined thus: the point (x_1, x_2, x_3) is said to belong to the line (u_1, u_2, u_3) if $u_1x_1 + u_2x_2 + u_3x_3 = 0$. The computations leading to formulas (β) imply that this condition has an invariant meaning (that is, if this condition is fulfilled for the given real or imaginary point and the given real or imaginary line in one coordinate system, then it is also fulfilled in any other coordinate system).

For imaginary objects, order relationships are not introduced. The set of real points and lines of the projective plane augmented

by imaginary elements will be referred to as the *complex projective plane*.

Like points and lines, all the remaining algebraic images of the complex projective plane are also divided into real and imaginary images. Those algebraic images which can be represented by equations with real coefficients are real, while imaginary images can be represented only by equations with complex coefficients. Lest one forms a mistaken impression, we immediately note that there may be real images which are composed entirely of imaginary elements; for example, the curve $x_1^2 + x_2^2 + x_3^2 = 0$ is real although it does not have any real point.*

On the complex projective plane every algebraic curve of order n has n points of intersection with any line (having regard to the multiplicity of the points). To see this, let us return to the considerations of the beginning of the preceding article. Equation (*) has n real or complex roots $\lambda_1, \lambda_2, \dots, \lambda_n$ with each one of which there are associated three numbers according to formulas

$$x_i = p_i + \lambda q_i \quad (i = 1, 2, 3)$$

Since imaginary elements have been introduced, the three numbers (x_1, x_2, x_3) can be treated, irrespective of whether they are real or complex, as the coordinates of some point. The points corresponding to the roots $\lambda_1, \lambda_2, \dots, \lambda_n$ are the common points of the curve and the line in question. While counting these points, if the points corresponding to the repeated roots are counted as many times, then there will always be n points of intersection of the line with a given n th order curve.

Similar arguments apply to pencils.

Thus on the projective plane

the order of an algebraic curve	the class of an algebraic pencil
equals the number of points on	equals the number of lines of
this curve that belong to	this pencil that pass through
a given line.	a given point.

We conclude this subsection with the remark that an algebraic pencil is, in general, composed of lines that are tangent to an algebraic curve.

Indeed, since every line $u_1x_1 + u_2x_2 + u_3x_3 = 0$ in the projective plane is determined by two parameters $u_1 \div u_2 \div u_3$ and the equation of a pencil establishes a connection between these parameters, it follows that an algebraic pencil is a one-parameter family of lines,

* If we classify all the transformations of coordinates according to formulas (α) and (β) with complex c_{ik} , then the distinction between imaginary and real images loses its meaning.

but a one-parameter family of lines has, in general, an envelope. Thus an algebraic pencil is composed of lines which are tangent to the curve. That this curve is algebraic is established by easy computations according to the general methods of the theory of envelopes.

The above remark will enable the reader to form a clear view of the algebraic pencils.

§ 128. The algebraic curves and algebraic pencils are evidently dual concepts of projective geometry. According to the principle of duality, a pencil of tangent lines to algebraic curves corresponds to the envelope of an algebraic pencil. In order to see this association, one has to take into account the fact that the envelope of a pencil consists of characteristic point each of which is a common point of two infinitely close lines of the pencil; it is clear that, according to the duality principle, a tangent line to the curve corresponds to a characteristic point, that is, the line passing through two infinitely close points of the curve. Consequently, to the set of characteristic points of an algebraic pencil (that is, to the envelope) there corresponds, as a dual image, the set of lines tangent to the algebraic curve (that is, the algebraic pencil enveloped by this curve).

In projective geometry, one often speaks of the class of a curve and the order of a pencil.

By the *class of an algebraic curve* is meant the class of algebraic pencil of its tangent lines. In other words, the class of a curve is the number of tangent lines (real or imaginary) that can be drawn to it from any point of the plane.

By the *order of an algebraic pencil* is meant the order of its envelope. In other words, the order of a pencil is the number of its characteristic points (real or imaginary) situated on one line.

The class and order of one and the same algebraic image are, in general, different.

§ 129. We shall not present here deeper theorems of the general theory of algebraic curves but confine ourselves only to a few remarks. The basic propositions of algebra and analysis imply that an algebraic curve, in contrast to some transcendental curves (that is, not algebraic curves), cannot have cessation points and cannot have the form of an infinite thread wrapped around the projective plane. In other words, all the algebraic curves are closed. For example, the well known algebraic curves in the Euclidean plane, the parabola and the hyperbola, are closed at infinity on the Euclidean plane augmented by elements at infinity, and thus are closed curves on the extended (that is, projective) plane.

Similarly, it may be proved that the number of separate pieces of every algebraic curve is finite.

As to the problem of classifying algebraic curves, this problem for $n > 3$ leads to the complicated reign of algebra (namely, to the invariant theory of homogeneous forms of three arguments) and this topic requires a special investigation.

§ 130. The imaginary elements may be adjoined to a real projective space in complete analogy with the plane case. Namely, one has to first define imaginary points and imaginary planes and the incidence relationship between the real and imaginary points and planes (analogous to the definitions of imaginary points, imaginary lines and incidence relationship given in § 127); then for an arbitrary line one has to treat the set of points of intersection of two given planes (and the new, that is, imaginary, lines are those which are not determined by the intersection of real planes). The set of real and imaginary elements thus obtained with the given incidence and order relationships (of real points on real lines) is called the *complex projective space*.

In the complex projective space one defines algebraic surfaces and bundles (which are the space analogues of the algebraic curves and pencils).

An *algebraic surface* is the set of points whose coordinates satisfy the equation

$$\sum a_{\alpha_1 \alpha_2 \dots \alpha_m} x_{\alpha_1} x_{\alpha_2} \dots x_{\alpha_m} = 0$$

$$(\alpha_1, \alpha_2, \dots, \alpha_m = 1, 2, 3, 4)$$

where the left-hand side is a homogeneous form of degree m in the variables x_1, x_2, x_3, x_4 . The number m is called the *order of the surface*.

An *algebraic bundle* is the set of planes whose coordinates satisfy the equation

$$\sum a_{\alpha_1 \alpha_2 \dots \alpha_m} u_{\alpha_1} u_{\alpha_2} \dots u_{\alpha_m} = 0$$

$$(\alpha_1, \alpha_2, \dots, \alpha_m = 1, 2, 3, 4)$$

where the left-hand side is a homogeneous form of degree m in the variables u_1, u_2, u_3, u_4 . The number m is called the *class of the bundle*.

Algebraic images are *real* if they can be represented by equations with real coefficients.

Extending the arguments advanced in § 126 to the case of three dimensions, we can prove that

the order of an algebraic surface equals the number of its points (real or imaginary) that belong to one line.

the class of an algebraic bundle equals the number of its planes (real or imaginary) that pass through one line.

In conclusion, we remark that not all the properties of equations of algebraic images express the geometrical properties of these images, but only those properties which remain unchanged under any transformation of projective coordinates.

Thus the problem of investigating algebraic surfaces and bundles in three-dimensional projective geometry is equivalent to the algebraic problem of investigating invariants of homogeneous forms containing four arguments.

5.13. Images of second degree. The theory of polars

The general problem of investigating algebraic images of a given order or class m consists in obtaining a complete system of invariants of a homogeneous equation of degree m , that is, in obtaining a system of functions of coefficients of the m th degree equation which

(1) are invariants under a homogeneous linear transformation of the arguments in the left-hand side of the equation;

(2) are such that if for two equations of degree m with numerical given coefficients these functions assume equal values, then the given equations are transformed into each other by means of a linear transformation of the arguments.

In other words, by knowing the complete system of invariants of an equation of degree m we can always settle the question whether any two arbitrary images of order or class m are projectively identical or not.

For large m this problem offers serious difficulties even in the two-dimensional case. For $m = 3$, this problem was advanced to a great extent by Newton*, who made a complete classification of 3rd order curves, that is, he found all the projectively distinct forms of these curves from which the remaining curves can be obtained by projective transformations. The case $m = 2$ is the simplest and can be investigated by elementary methods. This case will be dealt with in the present section in some detail.

We shall mainly confine ourselves to the case of two-dimensional geometry; almost all the results which will be obtained may be generalized to the case of three-dimensional geometry by the standard changes in the formulations and equations. Note further that in the investigation of 2nd degree images it suffices to deal with the curves of the 2nd order only, since the properties of pencils of the 2nd class** may be obtained by the duality principle.

We start by setting forth the polar theory, which plays an important role in the general investigation of images of the 2nd degree.

§ 131. Here we give the definition and main properties of polars. Consider a (real) conic given by the equation

$$\sum a_{ik}x_i x_k = 0 \quad (a_{ik} = a_{ki}) \quad (\alpha)$$

* See F. Klein, *Vorlesungen über höhere Geometrie*, Springer, Berlin, 1926 (Section 36).

** Curves of the 2nd order will henceforth be called conics and pencils of the 2nd class line conics.

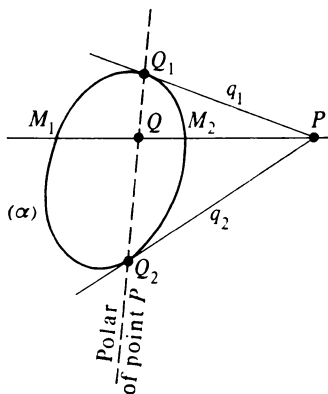


Figure 119

which in the expanded form is

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 + a_{33}x_3^2 = 0$$

We shall say that two points P and Q are situated harmonically with respect to the given conic (α) if the pair of points P, Q is harmonically conjugate to the pair of points M_1, M_2 where the conic intersects the line PQ (Fig. 119).

The geometrical figure consisting of points which together with P are situated harmonically with respect to a conic is called the *polar* of P with respect to the given conic.

We shall show now that the polar is a line. To this end, let us derive the equation of the polar.

First we shall obtain a condition on the coordinates p_i and q_i of P and Q for the points P, Q to be situated harmonically with respect to the conic (α) . We know that the coordinates x_i of any point M on the line PQ are of the form

$$x_i = p_i + \lambda q_i \quad (i = 1, 2, 3)$$

The common points of (α) and PQ will be obtained if for λ we choose the roots λ_1 and λ_2 of the quadratic equation

$$\sum a_{ik} (p_i + \lambda q_i) (p_k + \lambda q_k) = 0$$

which can be written as

$$\lambda^2 \sum a_{ik} q_i q_k + \lambda (\sum a_{ik} p_i q_k + \sum a_{ik} p_k q_i) + \sum a_{ik} p_i p_k = 0$$

or, since by symmetry $a_{ik} = a_{ki}$, in the form

$$\lambda^2 \sum a_{ik} q_i q_k + 2\lambda \sum a_{ik} p_i q_k + \sum a_{ik} p_i p_k = 0.$$

By § 119, two pairs of points p_i, q_i and $p_i + \lambda_1 q_i, p_i + \lambda_2 q_i$ are harmonically conjugate if $\lambda_1/\lambda_2 = -1$, that is, if $\lambda_1 + \lambda_2 = 0$. This together with Vietta's theorem yields the desired condition for the harmonic location of points P, Q with respect to the conic (α) :

$$\sum a_{ik} p_i q_k = 0 \quad (\beta)$$

Taking Q to be any point which together with P is harmonically situated with respect to (α) and changing its coordinates from q_1, q_2, q_3 to x_1, x_2, x_3 , we obtain the equation of the polar of P

$$\sum a_{ik} p_i x_k = 0 \quad (\gamma)$$

with current coordinates x_k . In the expanded form equation (γ) is

$$(a_{11}p_1 + a_{21}p_2 + a_{31}p_3)x_1 + (a_{12}p_1 + a_{22}p_2 + a_{32}p_3)x_2 + (a_{13}p_1 + a_{23}p_2 + a_{33}p_3)x_3 = 0 \quad (\delta)$$

which shows that this equation is of first degree, and, accordingly, the polar is indeed a line.

If one sets $\sum a_{ik} p_i p_k = \Phi(p_1, p_2, p_3)$, the equation of the polar can be written as

$$\frac{\partial \Phi}{\partial p_1} x_1 + \frac{\partial \Phi}{\partial p_2} x_2 + \frac{\partial \Phi}{\partial p_3} x_3 = 0$$

The form of this equation shows that it does not differ from the equation of a line in homogeneous coordinates that is tangent to the curve $\Phi(p_1, p_2, p_3) = 0$ at the point (p_1, p_2, p_3) ; this equation is well known in analysis and in differential geometry. Since the definition of a tangent line and the derivation of its equation given in analysis are based only on those properties of the curves that hold in projective geometry, we may assert the following

Theorem 49. *If a point P lies on a conic, then the polar of this point is a line which is tangent to the given curve at this point.*

Next, the following important theorem regarding the polars of arbitrary points must be noted.

Theorem 50. (The mutuality principle in the theory of polars). *If the polar of a point P passes through the point Q , then the polar of Q passes through P .*

This theorem is an immediate consequence of the equation of a polar. Indeed, if p_i are the coordinates of P , the polar of P has the equation

$$\sum a_{ik} p_i x_k = 0$$

and if q_i are the coordinates of Q , its polar has the equation

$$\sum a_{ik} q_i x_k = 0$$

Since by symmetry $a_{ik} = a_{ki}$, we have

$$\sum a_{ik} p_i q_k = \sum a_{ik} q_i p_k$$

Therefore the relation $\sum a_{ik} p_i q_k = 0$, showing that Q lies on the polar of P , implies the relation $\sum a_{ik} q_i p_k = 0$, which shows that P lies on the polar of Q . ■

Theorems 49 and 50 immediately imply

Theorem 51. *Lines that pass through a point P and are tangent to a conic have their points of contact of the polar of P (Fig. 119).*

Indeed, if q_1 is the tangent line and Q_1 its point of contact, then, according to Theorem 49, the line q_1 is the polar of Q_1 , and because q_1 passes through P , Theorem 50 implies that the polar of P passes through Q_1 , as required. ■

We note that a completely visual proof of Theorem 51 can be given by treating the tangent PQ_1 as the limiting position of the secant PM_2M_1 .

§ 132. If the line p is the polar of a point P , this point P is known as the *pole* of line p .

A natural question would be: does every line have a pole? In order to answer this question, compare the equation of any line

$$u_1 x_1 + u_2 x_2 + u_3 x_3 = 0 \quad (\varepsilon)$$

with equation (δ) of a polar. Evidently, the line (ε) will be the polar of some point if this equation can be put in the form of the equation of the polar, that is, if there are numbers p_1, p_2, p_3 such that

$$\left. \begin{aligned} a_{11}p_1 + a_{21}p_2 + a_{31}p_3 &= u_1 \\ a_{12}p_1 + a_{22}p_2 + a_{32}p_3 &= u_2 \\ a_{13}p_1 + a_{23}p_2 + a_{33}p_3 &= u_3 \end{aligned} \right\} \quad (\zeta)$$

then the point with coordinates p_1, p_2, p_3 will be the pole of line (ε). The system (ζ) will have a solution for any values of u_1, u_2, u_3 if and only if the determinant

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \neq 0$$

Therefore we may state the following proposition:

If a conic satisfies the condition $\Delta \neq 0$, then with respect to this conic every line has a pole.

The curves for which $\Delta = 0$ will be called *degenerate* (a visual description of degenerate curves will be given in § 134).

Another important feature needs attention. For any point (p_1, p_2, p_3) we may form the equation of the polar according to formu-

la (δ). But a definite equation is not always obtained; namely, the possibility that

$$a_{11}p_1 + a_{21}p_2 + a_{31}p_3 = 0 \quad (*)$$

$$a_{12}p_1 + a_{22}p_2 + a_{32}p_3 = 0$$

$$a_{13}p_1 + a_{23}p_2 + a_{33}p_3 = 0$$

is not ruled out. Without going into detail, we note that if p_1, p_2, p_3 satisfy equations (*), then

$$\begin{aligned} \sum a_{ih}p_i p_h &= (a_{11}p_1 + a_{21}p_2 + a_{31}p_3)p_1 + (a_{12}p_1 + a_{22}p_2 + \\ &\quad + a_{32}p_3)p_2 + (a_{13}p_1 + a_{23}p_2 + a_{33}p_3)p_3 = 0 \end{aligned}$$

and therefore the point (p_1, p_2, p_3) lies on a conic. Thus

Indeterminate polars exist only for those points that lie on a given conic.

What is more, because the system (*) is inconsistent (the solution $p_1 = p_2 = p_3 = 0$ is rejected) when $\Delta \neq 0$ one may assert that all the points have definite polars with regard to a nondegenerate conic.

§ 133. Consider a nondegenerate conic. With every point of the plane one may associate a definite line, its polar, and with every line a definite point, its pole. It is easy to show that then

- (1) to distinct points there correspond distinct lines;
- (2) to distinct lines there correspond distinct points;
- (3) to the point of intersection of two lines there corresponds a line joining the poles of the given lines (this follows from Theorem 50);
- (4) to a line joining two points there corresponds the point of intersection of the polars of the given points (this also follows from Theorem 50).

In general, under the above correspondence of geometric elements to every figure A composed of points and lines there corresponds some figure A' which is known as the *polar transform* of A with respect to the given conic.

If figure A' is the polar transform of a figure A , then A , in turn, is the polar transform of A' ; therefore two such figures are known as *mutually polar*. A figure which coincides with its polar transform is called *selfpolar*. For instance, if we take any point P and take any point Q on the polar p of P , and let q denote the polar of Q and R the point of intersection of the lines p and q and let r denote the polar of R (Fig. 120), then, by Theorem 50, q will pass through P , and r will pass through P and Q ; thus we obtain a *selfpolar three-point* each side of which is the polar of the opposite vertex and hence each vertex is the pole of the opposite side. Considerable use of selfpolar

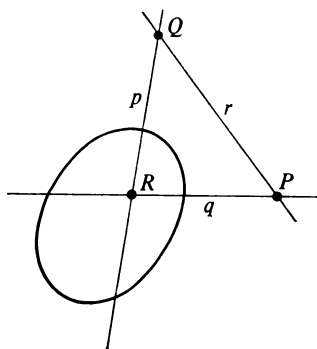


Figure 120

three-points will be made in the next article. Note also that mutually polar figures are at the same time mutually dual. The principle of duality was discovered by its author, Poncelet by the very polar transforms of figures.

§ 134. Now we are closer to solving the problem of determining all projectively distinct conics and of obtaining a complete system of invariants of 2nd degree equations. In order to find all the forms of conics we shall construct a system of coordinates in which the equation of any given conic assumes the simplest possible form.

Let a conic k be given. Take a point P outside this curve and denote its polar by p ; in view of the remark made at the end of § 132, the line p is completely defined because P does not lie on the curve. Introduce a projective coordinate system by taking the vertices of the coordinate trihedron in such a way that one vertex $A_1(1, 0, 0)$ coincides with P and the other two, $A_2(0, 1, 0)$ and $A_3(0, 0, 1)$ lie on p in any manner. The unit point $E(1, 1, 1)$ is chosen arbitrarily. Let

$$\sum a_{ik}x_i x_k = 0$$

be the equation of the conic k in the coordinate system just introduced. Now note that p has the equation

$$x_1 = 0 \quad (*)$$

because p is the side A_2A_3 of the coordinate trihedron. On the other hand, the equation of the same line as the polar of the point $A_1(1, 0, 0)$ may be formed according to formula (δ), § 131; substituting $p_1 = 1, p_2 = 0, p_3 = 0$ in this formula, we have

$$a_{11}x_1 + a_{12}x_2 + a_{13}x_3 = 0 \quad (**)$$

Since both equations (*) and (**) represent the same line, we must have

$$a_{12} = 0, \quad a_{13} = 0$$

Thus the equation of k in the present coordinate system is

$$a_{11}x_1^2 + a_{22}x_2^2 + 2a_{23}x_2x_3 + a_{33}x_3^2 = 0$$

If $a_{23} \neq 0$, we can further specialize the choice of the coordinate trihedron. Namely, the point A_2 is chosen on p in any manner subject to the sole condition that this point does not lie on k ; this is possible because when $a_{23} \neq 0$ there is a point $(0, x_2, x_3)$ on the line $x_1 = 0$ for which $a_{22}x_2^2 + 2a_{23}x_2x_3 + a_{33}x_3^2 \neq 0$. For A_3 , however, take the point of intersection of p and the polar of A_2 ; the choice of A_3 is no longer arbitrary because A_2 has a definite polar since it does not lie on k .

The coordinate trihedron $A_1A_2A_3$ thus obtained is selfpolar with respect to k , that is, each side of the trihedron is the polar of the opposite vertex. In particular, the line A_1A_3 whose equation is

$$x_2 = 0 \quad (***)$$

is the polar of the point $A_2(0, 1, 0)$; substituting $p_1 = 0, p_2 = 1, p_3 = 0$ in formula (δ) of § 131, we obtain the equation of the polar of the point A_3 in the form

$$a_{21}x_1 + a_{22}x_2 + a_{23}x_3 = 0$$

or, since $a_{21} = a_{12} = 0$,

$$a_{22}x_2 + a_{23}x_3 = 0 \quad (****)$$

Comparing equations (***) and (****),

$$a_{23} = 0$$

Therefore by a suitable choice of the coordinate trihedron the equation of a conic can always be reduced to the form

$$a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 = 0 \quad (\text{I})$$

In order to simplify this equation further, we must distinguish three cases:

(1) If $a_{22} = a_{33} = 0$, then equation (I) has the form

$$x_1^2 = 0 \quad (1)$$

and further simplification is impossible.

(2) If $a_{33} = 0$ and $a_{11} \neq 0, a_{22} \neq 0$, then by a transformation of coordinates*

$$x'_1 = x_1 \sqrt{|a_{11}|}, \quad x'_2 = x_2 \sqrt{|a_{22}|}, \quad x'_3 = x_3$$

* We once again recall that only real curves and real transformations are considered (that is, all the coefficients in the equations of the curves and in formulas of transformations are assumed to be real).

equation (I) may be reduced to the form

$$\pm x_1'^2 \pm x_2'^2 = 0 \quad (2)$$

(3) If $a_{11} \neq 0$, $a_{22} \neq 0$, $a_{33} \neq 0$, then from the transformation

$$x_1' = x_1 \sqrt{|a_{11}|}, \quad x_2' = x_2 \sqrt{|a_{22}|}, \quad x_3' = x_3 \sqrt{|a_{33}|}$$

we obtain

$$\pm x_1^2 \pm x_2^2 \pm x_3^2 = 0 \quad (3)$$

These simplifications, which are made after a coordinate trihedron $A_1A_2A_3$ has been chosen, obviously require changes in the unit points.

The simple equations (1), (2) and (3) of a conic are known as the *canonical equations*. By a suitable redesignation of the coordinates and multiplication by -1 these equations may be reduced to the following:

$$x_1^2 = 0 \quad (1)$$

$$\left. \begin{aligned} x_1^2 + x_2^2 &= 0 \\ x_1^2 - x_2^2 &= 0 \end{aligned} \right\} \quad (2)$$

$$\left. \begin{aligned} x_1^2 + x_2^2 + x_3^2 &= 0 \\ x_1^2 + x_2^2 - x_3^2 &= 0 \end{aligned} \right\} \quad (3)$$

Equation (1) defines the line $x_1 = 0$ taken twice. Each of the equations (2) defines a pair of distinct lines; namely, the equation $x_1^2 + x_2^2 = 0$ defines the pair $x_1 + ix_2 = 0$ and $x_1 - ix_2 = 0$ of imaginary lines, whereas the equation $x_1^2 - x_2^2 = 0$ the pair $x_1 + x_2 = 0$ and $x_1 - x_2 = 0$ of real lines.

All the conics (1) and (2) are degenerate because for equations $x_1^2 = 0$, $x_1^2 + x_2^2 = 0$ and $x_1^2 - x_2^2 = 0$ we have, respectively,

$$\Delta = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} = 0, \quad \Delta = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{vmatrix} = 0,$$

$$\Delta = \begin{vmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{vmatrix} = 0$$

Equations (3) define nondegenerate conics because for these equations

$$\Delta = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \pm 1 \end{vmatrix} \neq 0$$

The degenerate curves are therefore pairs of lines. The first of equations (3) defines a curve that does not have any real points, and this curve is called a *null curve*. The second of equations (3) defines a curve in the exact sense of the word, and it is called an *oval curve*. The oval curve divides the (real) projective plane into two regions, the first of which is characterized by the condition

$$x_1^2 + x_2^2 - x_3^2 < 0$$

and is known as the *interior region*, and the second by the condition

$$x_1^2 + x_2^2 - x_3^2 > 0$$

and is known as the *exterior region*. In order to clearly see the structure of these regions, note that the line $x_3 = 0$ does not intersect the interior region because when $x_3 = 0$ and x_1, x_2 are real the inequality $x_1^2 + x_2^2 - x_3^2 < 0$ is impossible; therefore for all the points of the interior region $x_3 \neq 0$ and we may determine these points by non-homogeneous coordinates $x = x_1/x_3, y = x_2/x_3$. In the nonhomogeneous coordinates the interior region is characterized by the condition

$$x^2 + y^2 < 1$$

and is, accordingly, topologically equivalent to the Euclidean disc*: this implies that the exterior region is a Möbius strip (see § 240).

Now we are in a position to obtain a complete system of the invariants of the general equation

$$\sum a_{ik} x_i x_k = 0$$

of a conic.

First of all one of the invariants of this equation is the rank of its matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Here we shall refer to the well known proposition of algebra (in the chapter on quadratic forms) stating that if the arguments x_i of the quadratic form $\sum a_{ik} x_i x_k$ are replaced by a homogeneous linear functions of the new arguments x'_i , then, provided the determinant of the coefficients of these functions does not vanish, the quadratic form $\sum a'_{ik} x'_i x'_k$ resulting from such a transformation has a matrix A' whose rank is the same as that of the matrix A of the original form:

$$\text{Rank } A' = \text{Rank } A$$

* Two figures are *topologically equivalent* if the point set of one of the figures may be mapped into the point set of the other by a mapping which is one-to-one and continuous in both directions. For example, a square and a disc are topologically equivalent; and so are a cube and a ball. In contrast, a ball and a torus are topologically distinct.

An examination of equations (1), (2) and (3) reveals that the rank of the matrix of equation (1) is 1, the rank of the matrix of equation (2) is 2 and that of equation (3) is 3.

Since the rank of a matrix is invariant, by looking at the rank of the matrix of the equation of any conic in any coordinate system it is possible to determine to which of the three canonical forms (1), (2) and (3) this equation may be reduced.

Secondly, an invariant of the equation $\sum a_{ik}x_ix_k=0$ is the signature of the equation.

The *signature* of a quadratic form is the absolute value of the difference of the number of positive and negative members in the canonical form of the given equation. The invariance of the signature expresses the theorem on inertia of quadratic forms known from algebra: all canonical representations of a quadratic form obtained by means of various real linear transformations have the same signature.

By knowing the signature of the left-hand side of the general equation of the conic, besides the rank of matrix A , not only are we able to point out to which of the three groups (1), (2), (3) of canonical equations the given equation may be reduced but also to which equation inside the particular group.

Thus the rank of the matrix and the signature of the equation of a conic constitute a complete system of its invariants.

We find that the equation of a conic has only two invariants for the integral values of which there are only five different combinations; accordingly, there are only five projectively different conics and all the remaining curves may be obtained from them by projective transformations.

The classification of conics are given in the following table:

Canonical form of the equation	Form of the curve	Rank	Signature
$x_1^2=0$	Pair of coincident lines	1	1
$x_1^2+x_2^2=0$	Pair of imaginary lines	2	2
$x_1^2-x_2^2=0$	Pair of real lines	2	0
$x_1^2+x_2^2+x_3^2=0$	Null curve	3	3
$x_1^2+x_2^2-x_3^2=0$	Oval curve	3	1

In contrast to conics, the curves of higher orders have always continual invariants and even for the class of 3rd order curves there is an infinite set of projectively distinct curves.

§ 135. We shall now briefly set forth the main facts from the theory of line conics. The general equation of the line conic is

$$\sum a_{ik} u_i u_k = 0 \quad (\alpha)$$

or, in the expanded form,

$$a_{11}u_1^2 + 2a_{12}u_1u_2 + a_{22}u_2^2 + 2a_{13}u_1u_3 + 2a_{23}u_2u_3 + a_{33}u_3^2 = 0$$

(here u_1, u_2, u_3 are the current coordinates of any line of the pencil). We shall say that the lines s and t are harmonically situated with respect to the given line conic (α) if the pair of lines s and t is harmonically conjugate to the pair of lines of the line conic (α) that pass through the common point of s and t .

In terms of the coordinates s_i, t_i of the lines s, t the condition for their harmonic location with respect to the line conic (α) may be expressed in the form

$$\sum a_{ik} s_i t_k = 0 \quad (\beta)$$

This condition follows from the dual derivation of condition (β) in § 131.

This implies that the set of lines that are harmonically situated together with a fixed line s with respect to the line conic (α) is determined by the equation

$$\sum a_{ik} s_i u_k = 0 \quad (\gamma)$$

where u_k are the current coordinates (that is, the coordinates of any line of the set in question).

The equation (γ) is of first degree; therefore the lines which are harmonically situated together with a fixed line s constitute a pencil of the first class, the centre S of which is called the *pole* of s with respect to the given line conic.

If we set $\sum a_{ik} s_i s_k = \Phi(s_1, s_2, s_3)$, equation (γ) may be written in the form

$$\frac{\partial \Phi}{\partial s_1} u_1 + \frac{\partial \Phi}{\partial s_2} u_2 + \frac{\partial \Phi}{\partial s_3} u_3 = 0 \quad (\delta)$$

the coefficients $\partial \Phi / \partial s_1, \partial \Phi / \partial s_2, \partial \Phi / \partial s_3$ of equation (δ) are the coordinates of S .

The pole of a line with respect to the line conic is an image which is the dual to the polar of a point with respect to the conic. The line conic (α) is a family of lines defined by the equations

$$u_1 x_1 + u_2 x_2 + u_3 x_3 = 0 \quad (*)$$

where the coefficients u_k are connected by the relation

$$\Phi(u_1, u_2, u_3) = 0$$

We shall find characteristic points of this family, that is, the points of contact of the lines of the family with the envelope. According to

differential geometry, the characteristic point of a line (u_1, u_2, u_3) is determined by equation (*) with the additional condition

$$x_1 du_1 + x_2 du_2 + x_3 du_3 = 0 \quad (**)$$

where du_1, du_2, du_3 are connected by the relation

$$\frac{\partial \Phi}{\partial u_1} du_1 + \frac{\partial \Phi}{\partial u_2} du_2 + \frac{\partial \Phi}{\partial u_3} du_3 = 0 \quad (***)$$

Comparing the equations (**) and (***) and taking note of the fact that there are no conditions imposed on the quantities du_1, du_2, du_3 , apart from (***), we may conclude that x_1, x_2, x_3 are proportional to the quantities $\partial \Phi / \partial u_1, \partial \Phi / \partial u_2, \partial \Phi / \partial u_3$. In other words, $\partial \Phi / \partial u_1, \partial \Phi / \partial u_2, \partial \Phi / \partial u_3$ are the coordinates of the characteristic point of the line (u_1, u_2, u_3) .

We noted earlier in § 131 that the coordinates of the tangent at the point (p_1, p_2, p_3) of the conic $\Phi(p_1, p_2, p_3) = 0$ are the numbers $\partial \Phi / \partial p_1, \partial \Phi / \partial p_2, \partial \Phi / \partial p_3$. We find that the coordinates of characteristic points of a line conic are determined by the equation of this line conic in exactly the same way as the coordinates of the tangent lines to a conic are determined by the equation of this conic. This is due to the fact that the characteristic points of a line conic are images dual to the tangent lines to a conic.

We further note that the expressions $\partial \Phi / \partial s_1, \partial \Phi / \partial s_2, \partial \Phi / \partial s_3$ for the coordinates of the pole of the line (s_1, s_2, s_3) do not differ in form from the expressions for the coordinates of a characteristic point. This implies a theorem which is dual to Theorem 49:

If the line s belongs to a line conic, its pole is a characteristic point.

Just like conics, the line conics are classified into degenerate and nondegenerate line conics. The degenerate line conics are characterized by the condition $\Delta = 0$, where Δ is the third order determinant composed of the coefficients of the equation of the line conic. By knowing the geometrical meaning of the degeneracy of conics, we can easily know the geometrical meaning of degeneracy of line conics by the method of duality: since a degenerate conic is the set of points belonging to either of the two determined lines (Fig. 121a), a degenerate line conic is the set of lines belonging to either of the two determined points (Fig. 121b). In other words, a degenerate line conic is a pair of pencils of the first class (which may be distinct or coincident, real or imaginary).

As regards nondegenerate line conics, they have a simple connection with the nondegenerate conics, which is expressed by

Theorem 52. *The set of tangent lines to a nondegenerate conic is the nondegenerate line conic; the envelope of a nondegenerate line conic is a nondegenerate conic.*

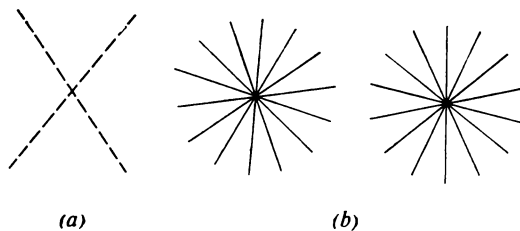


Figure 121

Proof. We shall prove the first part of this theorem, and the second part will then follow from the principle of duality. Let

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 + a_{33}x_3^2 = 0$$

be the equation of a conic k , and let p_1, p_2, p_3 be the coordinates of the point of contact with the line $u_1x_1 + u_2x_2 + u_3x_3 = 0$. A comparison of the general equation of the tangent

$$(a_{11}p_1 + a_{21}p_2 + a_{31}p_3)x_1 + (a_{12}p_1 + a_{22}p_2 + a_{32}p_3)x_2 + (a_{13}p_1 + a_{23}p_2 + a_{33}p_3)x_3 = 0$$

with the equation $u_1x_1 + u_2x_2 + u_3x_3 = 0$ gives the equations

$$\begin{aligned} a_{11}p_1 + a_{21}p_2 + a_{31}p_3 &= \alpha u_1 \\ a_{12}p_1 + a_{22}p_2 + a_{32}p_3 &= \alpha u_2 \\ a_{13}p_1 + a_{23}p_2 + a_{33}p_3 &= \alpha u_3 \end{aligned} \quad (*)$$

where $\alpha (\neq 0)$ is a proportionality factor. What is more, since the point of contact lies on the tangent line, we have

$$u_1p_1 + u_2p_2 + u_3p_3 = 0 \quad (**)$$

If the quantities p_1, p_2, p_3 are obtained from equations (*) (which is possible because $\Delta \neq 0$) and then substituted in (**), we shall obtain a relation connecting u_1, u_2, u_3 that can be treated as the condition for the contact of the line having coordinates (u_1, u_2, u_3) with the given conic k . The same relation may be obtained by equating to zero the determinant of the system composed of equations (*) and (**):

$$\begin{vmatrix} a_{11} & a_{21} & a_{31} & u_1 \\ a_{12} & a_{22} & a_{32} & u_2 \\ a_{13} & a_{23} & a_{33} & u_3 \\ u_1 & u_2 & u_3 & 0 \end{vmatrix} = 0 \quad (***)$$

The condition (***) characterizes the coordinates of the tangent. Therefore, as an equation with current coordinates u_1, u_2, u_3 it determines a pencil of tangent lines to the conic k .

We see that (***) is an equation of the second degree. This means that the tangent lines to a nondegenerate conic indeed constitute a line conic.

We have to show further that this line conic is nondegenerate.

To this end, note that if the left-hand side of equation (***) is expanded, we obtain a quadratic form

$$A_{11}u_1^2 + 2A_{12}u_1u_2 + A_{22}u_2^2 + 2A_{13}u_1u_3 + 2A_{23}u_2u_3 + A_{33}u_3^2 = 0$$

whose coefficients A_{ik} are the minors of a_{ik} in the matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

Therefore, if we divide the left-hand side of the equation of the line conic by Δ and then set $A_{ik}/\Delta = a_{ik}^*$, the equation of the line conic assumes the form

$$a_{11}^*u_1^2 + 2a_{12}^*u_1u_2 + a_{22}^*u_2^2 + 2a_{13}^*u_1u_3 + 2a_{23}^*u_2u_3 + a_{33}^*u_3^2 = 0$$

and the matrix A^* of this equation is the inverse of matrix A . But then, as we know, the determinants Δ and Δ^* of A and A^* are connected by the relation $\Delta\Delta^* = 1$, which implies that $\Delta^* \neq 0$, as required.

The above theorem may also be stated thus.

For every nondegenerate image of the second degree the class and order coincide (= 2).

Such an assertion is not valid for images of higher orders.

Since nondegenerate conics are also of the second class, it follows that through every point of the plane two tangent lines (distinct or coincident, real or imaginary) may be drawn to a conic.

§ 136. The methods just described for investigating the images of the second degree for the case of two-dimensional geometry can be extended in a natural manner to the three-dimensional case and lead to similar results. Namely, in the projective space, like in the projective plane, the images of the second degree are completely characterized by the integral values of only two invariants: the rank of the principal matrix and the signature of the left-hand side of the equation.

In the projective space there are only a finite number of projectively distinct surfaces of the second order and bundles of the second class*, and all the remaining surfaces and bundles can be obtained from them by projective transformations.

* Surfaces of the second order and bundles of the second class will be called quadrics and plane-quadrics, respectively.

For example, depending on the rank and signature of the equation

$$a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + a_{44}x_4^2 + 2a_{12}x_1x_2 + 2a_{13}x_1x_3 + 2a_{14}x_1x_4 + 2a_{23}x_2x_3 + 2a_{24}x_2x_4 + 2a_{34}x_3x_4 = 0$$

every quadric surface may be projectively transformed into one of the surfaces given in the following table:

Rank 1	Signature	Rank 2	Signature	Rank 3	Signature	Rank 4	Signature
$x_1^2 = 0$	1	$x_1^2 + x_2^2 = 0$ $x_1^2 - x_2^2 = 0$	2 0	$x_1^2 + x_2^2 + x_3^2 = 0$ $x_1^2 + x_2^2 - x_3^2 = 0$	3 1	$x_1^2 + x_2^2 + x_3^2 + x_4^2 = 0$ $x_1^2 + x_2^2 + x_3^2 - x_4^2 = 0$ $x_1^2 + x_2^2 - x_3^2 - x_4^2 = 0$	4 2 0

The equation $x_1^2 = 0$ defines a pair of coincident planes $x_1 = 0$. Each of the equations $x_1^2 + x_2^2 = 0$, $x_1^2 - x_2^2 = 0$ defines a pair of planes with the equation $x_1^2 + x_2^2 = 0$ defining a pair of imaginary planes $x_1 + ix_2 = 0$ and $x_1 - ix_2 = 0$ and the equation $x_1^2 - x_2^2 = 0$ a pair of real planes $x_1 + x_2 = 0$ and $x_1 - x_2 = 0$.

Each of the equations $x_1^2 + x_2^2 + x_3^2 = 0$ and $x_1^2 + x_2^2 - x_3^2 = 0$ defines a cone with the vertex at the point $A_4(0, 0, 0, 1)$, that is, a surface constituted by lines passing through $A_4(0, 0, 0, 1)$. In fact, if a point $M_0(x_1^0, x_2^0, x_3^0, x_4^0)$ lies, for example, on $\bar{x}_1^2 + \bar{x}_2^2 - \bar{x}_3^2 = 0$, then $\bar{x}_1^{02} + \bar{x}_2^{02} - \bar{x}_3^{02} = 0$; but the coordinates $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4$ of any point $\bar{M}$ of the line A_4M_0 satisfy $\bar{x}_1^2 + \bar{x}_2^2 - \bar{x}_3^2 = 0$, which can be easily shown by expressing the coordinates x_i parametrically:

$$\begin{aligned}\bar{x}_1 &= 0 + \lambda x_1^0 = \lambda x_1^0, & \bar{x}_3 &= 0 + \lambda x_3^0 = \lambda x_3^0 \\ \bar{x}_2 &= 0 + \lambda x_2^0 = \lambda x_2^0, & \bar{x}_4 &= 1 + \lambda x_4^0\end{aligned}$$

Indeed, we find that

$$\bar{x}_1^2 + \bar{x}_2^2 - \bar{x}_3^2 = \lambda^2 (x_1^{02} + x_2^{02} - x_3^{02}) = 0$$

Thus every point $\bar{M}$ of the line A_4M_0 lies on the surface $\bar{x}_1^2 + \bar{x}_2^2 - \bar{x}_3^2 = 0$, that is, the line joining A_4 with any point of the surface lies wholly on this surface, and this is the characteristic feature of the cone with vertex A_4 . Evidently, the cone $x_1^2 + x_2^2 + x_3^2 = 0$ does not have real generators, and this cone is referred to as the *null cone*. The cone $x_1^2 + x_2^2 - x_3^2 = 0$ has an infinite number of real generators and is referred to as the *ordinary cone*.

All the surfaces just mentioned are known as *degenerate*, a criterion for the degeneracy of a surface is the vanishing of the determinant of the matrix of the equation.

Nondegenerate quadric surfaces are given in the last column of the aforementioned table.

The first of them does not contain any real point and is called a *null surface*.

The second is known as the *oval surface*, this is topologically equivalent to the Euclidean sphere. In order to demonstrate this, first note that for all points on the surface $x_1^2 + x_2^2 + x_3^2 - x_4^2 = 0$ necessarily $x_4 \neq 0$; therefore these points can be determined by the nonhomogeneous coordinates $x = x_1/x_4$, $y = x_2/x_4$, $z = x_3/x_4$; but in nonhomogeneous coordinates the surface in question is of the form $x^2 + y^2 + z^2 = 1$ and is the same as the equation of the Euclidean sphere.

The third surface is called an *annular surface*, topologically it is equivalent to a torus. This is easily understood if one notices that the surface $x_1^2 + x_2^2 - x_3^2 - x_4^2 = 0$ is covered by lines. In fact, the two first degree equations

$$\left. \begin{aligned} \mu (x_1 + x_3) &= \lambda (x_2 + x_4) \\ -\lambda (x_1 - x_3) &= \mu (x_2 - x_4) \end{aligned} \right\} \quad (*)$$

define for any λ, μ (not both simultaneously zero) a line lying on the given surface since the equation $x_1^2 + x_2^2 - x_3^2 - x_4^2 = 0$ is a consequence of equation (*); further, through every point of the surface there passes one and only one line of the system (*) because for any $x_1^0, x_2^0, x_3^0, x_4^0$ satisfying $x_1^{02} + x_2^{02} - x_3^{02} - x_4^{02} = 0$ one can find only one ratio $\lambda_0 \div \mu_0$ such that

$$\left. \begin{aligned} \mu_0 (x_1^0 + x_3^0) &= \lambda_0 (x_2^0 + x_4^0) \\ -\lambda_0 (x_1^0 - x_3^0) &= \mu_0 (x_2^0 - x_4^0) \end{aligned} \right\}$$

Therefore for $\lambda = \lambda_0, \mu = \mu_0$ the line (*) passes through the given point $(x_1^0, x_2^0, x_3^0, x_4^0)$. Thus the real lines (*) entirely and singly cover the surface. But the system of these lines constitutes a closed "tube" because, on the one hand, the projective line is closed and, on the other hand, the lines of the system (*) pass through the oval curve, namely through the section of the surface $x_1^2 + x_2^2 - x_3^2 - x_4^2 = 0$ by the plane $x_4 = 0$.

Besides the system (*), the given surface is also covered by the system of lines defined by the equations

$$\left. \begin{aligned} \mu (x_1 + x_3) &= \lambda (x_2 - x_4) \\ -\lambda (x_1 - x_3) &= \mu (x_2 + x_4) \end{aligned} \right\} \quad (**)$$

The lines (*) and (**) are called the *rectilinear generators* of the surface.

Let us note that the hyperboloid of one sheet, a surface in the Euclidean space which is well known to the reader, becomes an annular surface of projective space if the elements at infinity are adjoined to the Euclidean space.

5.14. Constructive theorems and problems of projective geometry

§ 137. In the present section we shall introduce the reader to certain elementary theorems of projective geometry regarding conics. They have wide applications in the Euclidean problems on construction.

The reader might ask the question: Do we have a right to apply the theorems of projective geometry to investigation of figures in the Euclidean plane? In order to be convinced of the possibility of such an application it is enough to recall that by adjoining the elements at infinity the Euclidean plane turns into the projective plane. Thus every projective proposition can be interpreted on the Euclidean plane if this plane is assumed to have been augmented by the line at infinity.

As an example, consider from the point of view of elementary geometry the cross ratio of four points of a line and the cross ratio of four rays of a pencil.

Let A, B, C, D be any four points of a Euclidean line a . On the Euclidean plane containing line a introduce a Cartesian coordinate system. For convenience, we let the x -axis coincide with a , and denote the abscissas of the points A, B, C, D by x_a, x_b, x_c, x_d . Note that the Cartesian system on the Euclidean plane is simultaneously a projective system because all the lines in the Cartesian system are represented by first degree equations. Hence

$$(ABCD) = \frac{x_c - x_a}{x_b - x_c} \div \frac{x_d - x_a}{x_b - x_d}$$

(see § 115). But in the Cartesian coordinates, $x_c - x_a = AC$, $x_b - x_c = CB$, $x_d - x_a = AD$, $x_b - x_d = DB$, where AC, CB, AD, DB denote the signed lengths of the segments with the endpoints $A, C; C, B$, etc. Hence

$$(ABCD) = \frac{AC}{CB} \div \frac{AD}{DB} \quad (*)$$

The cross ratio of the four rays a, b, c, d that pass through a point O may be defined in elementary geometry by the formula

$$(abcd) = \frac{\sin(ac)}{\sin(cb)} \div \frac{\sin(ad)}{\sin(db)} \quad (**)$$

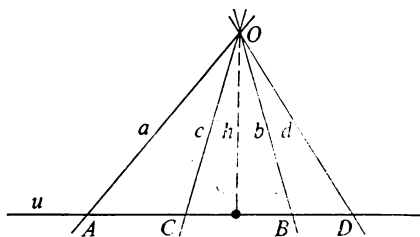


Figure 122

where (ac) , (cb) , etc. denote the signed angles between the rays a , c ; c , b , etc.*

In order to demonstrate the validity of formula (**), we intersect the rays a , b , c , d by a line u and denote the points of intersection by A , B , C , D and let h be the length of the perpendicular from O to the line u (Fig. 122). Expressing the area of $\triangle OAC$ in two ways, we have

$$\frac{1}{2} AC \times h = \frac{1}{2} OA \times OC \times \sin(ac)$$

Similarly, from $\triangle OCB$,

$$\frac{1}{2} CB \times h = \frac{1}{2} OB \times OC \times \sin(cb)$$

Therefore

$$\frac{AC}{CB} = \frac{OA}{OB} \frac{\sin(ac)}{\sin(cb)} \quad (1)$$

In the same manner, from the triangles OAD and ODB , we have

$$\frac{AD}{DB} = \frac{OA}{OB} \frac{\sin(ad)}{\sin(db)} \quad (2)$$

and from (1) and (2) we have

$$\frac{AC}{CB} \div \frac{AD}{DB} = \frac{\sin(ac)}{\sin(cb)} \div \frac{\sin(ad)}{\sin(db)}$$

Now it is clear that formula (**) is valid because the cross ratio $(abcd)$ is, by definition, the number $(ABCD)$ (see § 116).

* Namely, we choose the direction of positive rotation about the point O and positive directions on the lines a , b , c , d . Then, for example, (ab) denotes the magnitude of the angle which the positive direction of the line b makes with the positive direction of the line a ; this magnitude will be taken as positive if the rotation about O from a to b inside the angle is performed in the positive direction, and negative otherwise. It may be shown that the right-hand side of (**) does not depend on the choice of the positive rotations about O and the positive direction on each of the lines a , b , c , d .

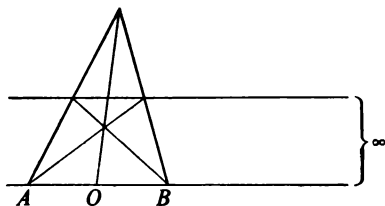


Figure 123

We have now the examples of metric interpretation of projective objects. In turn, the propositions of elementary geometry, even though having metric character, may be interpreted projectively and can often be looked at from a new angle if interpreted from the projective point of view.

For instance, the theorem of elementary geometry which states that the line joining the point of intersection of the diagonals of a trapezium with the point of intersection of the nonparallel sides divides the parallel sides into two halves has a clear projective interpretation, namely that the midpoint of the segment together with the point at infinity divides the pair of its endpoints harmonically. Indeed, looking at the trapezium drawn in Fig. 123, we can easily see a complete four-point with one diagonal point at infinity; from the geometrical properties of this four-point it follows that the pair A, B is harmonically divided by the pair $0, \infty$. In other words, the Euclidean centre of the segment coincides with its projective centre.

As noted somewhat earlier, every Cartesian system of coordinates on the Euclidean plane is at the same time a projective system. This implies that conics studied in analytic geometry are the same objects to the study of which the preceding sections were devoted. Precisely, they become the same objects after adjoining the elements at infinity to the Euclidean plane. Indeed, if starting from the Cartesian coordinates x, y we introduce homogeneous coordinates x_1, x_2, x_3 by setting $x = x_1/x_3, y = x_2/x_3$, then the general equation

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0$$

of a conic in the Cartesian coordinates assumes the form

$$a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2 + 2a_{13}x_1x_3 + 2a_{23}x_2x_3 + a_{33}x_3^2 = 0$$

and coincides with the equation studied earlier.

A number of propositions and problems are discussed below; all may be regarded as propositions and problems of elementary geometry.

§ 138. This subsection occupies an isolated position in the present section and is unconnected with the remaining material. Here

propositions of an auxiliary nature are described that will be used much later (in Chapter 9).

In the Euclidean plane endowed with a Cartesian coordinate system consider a disc k . Let there be given a one-to-one mapping of the interior of the disc into itself, and call this mapping φ . Assume that φ maps points lying on any one line into points also lying on one line and points not lying on one line into points also not lying on one line. In other words, it is assumed that φ maps every chord of k into a chord and distinct chords into distinct chords.

Under these conditions on φ the coordinates of the image point are expressed in terms of the coordinates of the preimage point by the formulas

$$x' = \frac{a_1x + b_1y + c_1}{\alpha x + \beta y + \gamma}, \quad y' = \frac{a_2x + b_2y + c_2}{\alpha x + \beta y + \gamma} \quad (1)$$

where $a_1, b_1, \dots, \gamma$ are constants satisfying the condition

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ \alpha & \beta & \gamma \end{vmatrix} \neq 0 \quad (2)$$

To prove this proposition, let P be any point in the interior of k and let a, b, c, d be four chords passing through P . Assume that a, b divide c, d harmonically. Then a four-point G may be constructed such that its vertices lie on the chords a, b and the diagonal points on the chords c, d . Under φ the point P is mapped into a point P' and the chords a, b, c, d are mapped into the chords a', b', c', d' , the four point G is mapped into a quadrangle G' whose vertices lie on the chords a', b' and the diagonal points on the chords c', d' .

This means that the chords a', b' divide the chords c', d' harmonically. We find that under φ any harmonic quadruple of chords is mapped into a harmonic quadruple of chords. Accordingly, every pencil of chords is projectively mapped into a corresponding pencil of chords.

We choose four points A, B, C, D inside k so that no three of them are collinear. Then, according to the conditions imposed on φ , no three of the four image points A', B', C', D' are collinear.

By Theorem 37, there is a projectivity of the entire plane (augmented by the points at infinity) into itself that maps the points A, B, C, D into the points A', B', C', D' . Denoting this mapping by ψ , we shall show that φ and ψ coincide in the interior of k , that is, that for any point inside k the image under φ coincides with the image under ψ .

Take any point M inside k and denote its image under φ and ψ by M' and M'' , respectively. Both φ and ψ map the pencil of chords with the centre A projectively into the pencil of chords with centre A' ,

and in the case of φ the chord AM is mapped into the chord $A'M'$ and in the case of ψ into chord $A'M''$. In both cases, the triple of distinct chords AB, AC, AD of the pencil with centre A is mapped into one and the same triple of distinct chords $A'B', A'C', A'D'$ of the pencil with centre A' . But in view of Theorem 18 the projectivity of pencils is uniquely determined by prescribing three pairs of corresponding elements. Therefore the chord $A'M'$ must coincide with the chord $A'M''$, that is, the points M', M'', A' are collinear. Similarly, it may be proved that the points M', M'', B' and M', M'', C' are collinear. But since A', B', C' are noncollinear, we may conclude from the preceding result that the points M' and M'' coincide. Thus, inside k the mapping φ coincides with a projectivity of the given plane (augmented by elements at infinity) into itself. According to § 112, the Cartesian coordinates of the image point and the preimage point under any projectivity are connected by formulas of form (1) subject to condition (2). Thus the assertion is proved. ■

The following proposition also follows from the preceding result.

Let M_1, M_2 be any two points in the interior of the disc k and M'_1, M'_2 be their images under φ . Let P, Q and P', Q' be the points where the lines M_1M_2 and $M'_1M'_2$ intersect k and are such that the ordering of points P, Q, M_1, M_2 on the line M_1M_2 is the same as the ordering of points P', Q', M'_1, M'_2 on the line $M'_1M'_2$. Then

$$(P'Q'M'_1M'_2) = (PQM_1M_2) \quad (3)$$

In fact, with the stated conditions the points P', Q', M'_1, M'_2 are images of the points P, Q, M_1, M_2 under the mapping ψ . And because ψ is a projectivity, the cross ratio of the points $PQ_1M_1M_2$ is the same as the cross ratio of the image points $P'Q'M'_1M'_2$.

§ 139. A criterion for the perspective correspondence. Now we shall only deal with projective correspondences between two lines and two pencils; all the objects under discussion are assumed to belong to one plane. Of special importance for us will be the so-called *per-*

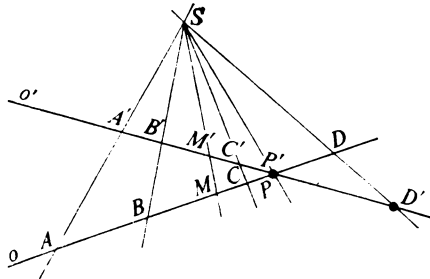


Figure 124

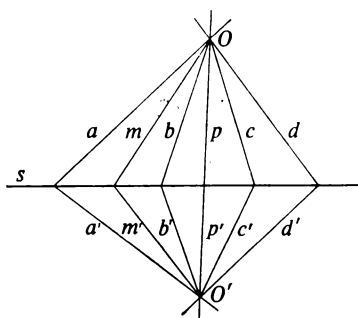


Figure 125

spective correspondences.* The correspondence between the points of two lines is called a *perspectivity* if the lines joining the corresponding points pass through one point of the plane (called the *centre of perspective*, Fig. 124).

The correspondence between the rays of two pencils is a *perspectivity* if the points of intersection of the corresponding rays lie on one line (called the *axis of perspective*; Fig. 125).

Obviously, every perspectivity is a projectivity (in view of Theorem 6 of § 86; see also § 103). However, not every projectivity is a perspectivity.

The following two theorems provide a useful criterion for a projectivity to be a perspectivity.

Theorem 53. *For a projectivity between points of two lines to be a perspectivity, it is necessary and sufficient that to the point of intersection of these lines, which is considered as an element of one of them, there corresponds the same point on the other line.*

Theorem 54. *For a projectivity between the rays of two pencils to be a perspectivity, it is necessary and sufficient that to the common ray of these pencils, which is considered as an element of one of them, there corresponds the same ray in the other pencil.*

Of these two theorems it suffices to prove only one; the other will follow by the duality principle. We shall prove Theorem 53.

That the condition is necessary is straightforward. In fact, if there is a perspectivity with the centre of perspective S between the lines o and o' , then the pairs of corresponding points are determined by the intersection of the lines o and o' with the rays originating from S ; but the ray passing through the common point of the lines o and o' evidently determines a pair of corresponding points P and P' which coincide with each other (Fig. 124).

Proof of sufficiency. Suppose that a projectivity is established between the points of the lines o and o' such that to a point P of o ,

* The term "perspectivity" is also used.

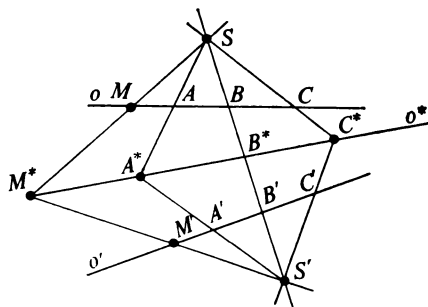


Figure 126

where the lines o and o' intersect, there corresponds a point P' coinciding with P on o' .

Take two points A, B on o and the corresponding points A', B' on o' , and let S denote the point of intersection of the lines AA', BB' . Furthermore, let M be any point on o and $M' = f(M)$ the corresponding point under the given mapping, and let $M^* = \varphi(M)$ be the point where the ray SM meets the line o' . By construction, the correspondence $M^* = \varphi(M)$ is a perspectivity (with centre of perspective S) and hence is also a projectivity; what is more, under the correspondence $M^* = \varphi(M)$ to the points A, B, P there evidently correspond the points A', B', P' . Thus we have the projectivities $M' = f(M)$ and $M^* = \varphi(M)$ with three common pairs of corresponding points $A, A'; B, B'$ and P, P' . According to Theorem 15, these correspondences cannot be distinct and hence $M' = M^*$. This implies that the given correspondence $M' = f(M)$ is a perspectivity with the centre of perspective S . ■

§ 140. Graphical construction of projectivities by means of three pairs of corresponding elements. Consider three points A, B, C on a line o and three points A', B', C' on another line o' . We know that there is a unique projectivity $M' = f(M)$ of o into o' that maps the points A, B, C into the points A', B', C' , respectively. Our aim at present is to indicate a graphical method for obtaining the point $M' = f(M)$ of o' that corresponds to the point M of o . To this end, join any two corresponding points, for example B and B' , among the given points by a line and on this line take two points S and S' (Fig. 126). Then we establish a correspondence between the rays of pencils with centres S and S' by assigning the ray $S'M'$ of the pencil S' to the ray SM of the pencil S . Since the correspondence $M' = f(M)$ is a projectivity, the correspondence established between the rays of pencils S and S' is also a projectivity. Moreover, it will also be a perspectivity because to the rays SB of the pencil S there corresponds the ray $S'B'$ of the pencil S' that coincides with SB

(see Theorem 54). Therefore all the corresponding rays of the pencils S and S' intersect on one line which is the axis of the perspective. Thus we have the desired construction: having chosen the points S and S' draw the rays SA , $S'A'$, SC , $S'C'$ and draw a line A^*C^* , as shown in Fig. 126. This will be the axis of perspective for the pencils S and S' .

In order to obtain a point M' on o' that corresponds to the point M on o , it is enough to draw the ray SM and join the point M^* , where this ray meets the line A^*C^* , with the points S' ; the line $S'M^*$ will intersect the line o' at the desired point $M' = f(M)$.

The process of constructing projectively corresponding rays of two pencils is the dual of the process just described. We shall present it without any detailed explanation.

Let O and O' be the centres of two pencils between the rays of which a projectivity has been established and let it be required to find a ray $m' = f(m)$ of the pencil O' corresponding to the ray m of the pencil O by knowing three rays a , b , c of the pencil O and three corresponding rays a' , b' , c' of pencil O' . To do this, draw through the point of intersection of any two corresponding rays, for example, through the point of intersection of b and b' , among the given corresponding rays, any two lines s and s' (Fig. 127). Then join by a line the point of intersection of a , s with the point of intersection of a' , s' and join the point of intersection of c , s by a line with the point of intersection of c' , s' ; the lines thus drawn are determined by their point of intersection O^* . Having done this, we construct the ray $m' = f(m)$ as shown in Fig. 127: the point of intersection of the ray m with the line s is joined with the point O^* and the point of intersection of this line with the line s' is determined; the ray originating from O' and going to this point is the desired ray $m' = f(m)$.

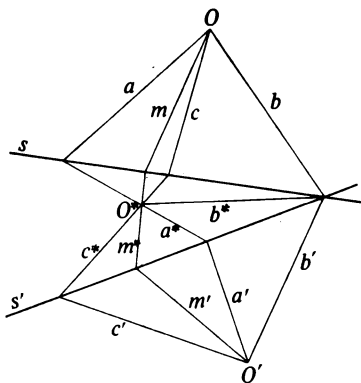


Figure 127

Note that the correspondence between the elements of one-dimensional manifolds established by means of a chain of operations of projection and section is always a projectivity (because these operations carry a harmonic set of elements into a harmonic set of elements). On the basis of the results obtained in this subsection we may assert that given any projectivity between the elements of one-dimensional manifolds this correspondence can always be obtained as a result of a chain of operations of projection and section.

§ 141. We now turn to the projective construction of images of the second degree (Steiner's theorem). The projectivity between the images of the first degree can be used to construct images of the second degree. The method of such a construction is contained in the following theorems, due to Steiner.

Theorem 55. *The set of points of intersection of projectively corresponding rays belonging to two pencils is a conic that passes through the centres of these pencils.*

Proof. Introduce a system of projective coordinates on the plane; the non-homogeneous system will be more convenient for the present case. Let $S_1(x_1, y_1)$ and $S_2(x_2, y_2)$ be the centres of two pencils between the rays of which a projectivity has been established. If

$$y - y_1 = k(x - x_1) \quad (1)$$

is the equation of any ray in the first pencil and

$$y - y_2 = k'(x - x_2) \quad (2)$$

is the equation of the corresponding ray belonging to the second pencil, we know that the parameter k' is a linear-fractional function of the parameter k :

$$k' = \frac{\alpha k + \beta}{\gamma k + \delta} \quad (\alpha\delta - \beta\gamma \neq 0) \quad (3)$$

(see § 119). In order to obtain the equation of the curve described by the point of intersection of the rays (1) and (2) for various values of k , one has to eliminate the parameters k and k' from the equations (1), (2), (3). The result is

$$\begin{aligned} \gamma(y - y_1)(y - y_2) + \delta(y - y_2)(x - x_1) \\ - \alpha(y - y_1)(x - x_2) - \beta(x - x_1)(x - x_2) = 0 \quad (*) \end{aligned}$$

which is obviously an equation of the second degree in x and y .

Thus the common points of the corresponding rays of the pencils S_1 and S_2 do constitute a conic. That this curve passes through S_1 and S_2 can be immediately demonstrated. In fact, both the values $x = x_1$, $y = y_1$ and $x = x_2$, $y = y_2$ satisfy equation (*), thereby meaning that the points S_1 and S_2 lie on the curve. By the way, the fact that the points S_1 and S_2 belong to the set of points of intersection of the corresponding rays of the pencils S_1 and S_2 is clear

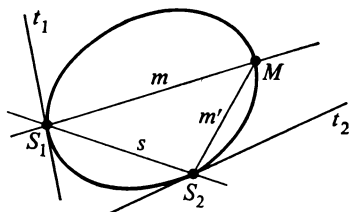


Figure 128

without any computations; for instance, S_2 belongs to this set because the ray S_1S_2 of the pencil intersects at S_2 all the rays of the second pencil, including the one that corresponds to the ray in question.

Note, by the way, that if the common ray S_1S_2 of the pencils S_1 and S_2 is assigned to the first pencil, then to this ray there corresponds in the second pencil a ray t_2 that touches at S_2 the conic formed by the projectively corresponding rays of pencils S_1 and S_2 ; if the ray S_1S_2 is assigned to the second pencil, then in the first pencil there is a corresponding ray t_1 that touches this curve at S_1 .

Indeed, let M be the point of intersection of two corresponding rays m and m' , s the common ray of pencils S_1 and S_2 , and t_2 the tangent at S_2 (Fig. 128). Assume that the point M tends to S_2 along the curve. Then m will tend to s and m' to t_2 . But a projectivity is continuous (this result follows from its analytical representation). Therefore the limiting positions of the corresponding rays $s = \lim_{M \rightarrow S_2} m$ and $t_2 = \lim_{M \rightarrow S_2} m'$ are also corresponding rays, that is, to ray s , considered as a ray of the first pencil, there corresponds a ray t_2 in the second pencil. Similarly, it can be shown that to s , considered as a ray of the second pencil, there corresponds a ray t_1 in the first pencil.

At this point one may ask the question as to whether any conic can be constructed by means of two projective pencils, that is, the pencils between the rays of which a projectivity has been established. Naturally, by means of projective pencils with real rays one can obtain only those conics which possess infinitely many real points. Therefore the following curves cannot be obtained: the degenerate conic composed of two imaginary lines and the nondegenerate null curve (this is clear because as soon as we remain within the limits of visual geometric constructions we lose sight of images composed of imaginary elements).

As regards the remaining conics, the pair of coincident real lines, the pair of distinct real lines and the oval curves (see § 134), all of them can be obtained in the manner described above.

(1) A pair of coincident real lines is a geometrical figure composed of the common points of corresponding rays of the projective pencils

if the centres of these pencils coincide and if a ray of one of the pencils which coincides with the corresponding ray of the second happens to be repeated. For example, the pencils

$$y = kx, \quad y = k'x \quad (*)$$

with a common centre $(0, 0)$ between whose rays a correspondence

$$k' = \frac{k}{1+k} \quad (**)$$

has been established have for the set of common points of the corresponding rays the coordinate axis taken twice because the elimination of the parameters k, k' from the equations $(*)$ and $(**)$ yields $y^2 = 0$.

(2) The pair of distinct real lines is formed by the intersection of the corresponding rays of the two pencils which are in a perspective correspondence. The axis of perspective is one line of the pair and the common ray of pencils the other line.

(3) The oval curve is formed by the intersection of the corresponding rays of two pencils when between these pencils there has been established a projectivity and not a perspectivity; indeed, in this case the common points of corresponding rays do not lie on lines.*

By the duality principle, from Theorem 55 one has

Theorem 56. *The set of lines joining the corresponding points of two fixed lines which are in a projective correspondence is a line-conic containing these two lines.*

If the fixed lines s_1 and s_2 are in a perspective correspondence, then the line-conic mentioned in Theorem 56 degenerates into a pair of pencils of the first class; one of them will have its centre at the centre of perspective of the given correspondence, and the other at the common point of the lines s_1 and s_2 .

If the lines s_1 and s_2 are in a projective but not in perspective correspondence, then the line-conic defined as in Theorem 56 will be nondegenerate. By Theorem 52, this pencil has an envelope which is a nondegenerate conic. And because the pencil contains s_1 and s_2 , the envelope, which is a conic, touches these lines.

Figure 129 shows a nondegenerate line-conic which is generated by the projectivity between the lines s_1 and s_2 ; the same figure also shows the envelope of this pencil which is a conic k (T_1 and T_2 denote the points of contact of k with s_1 and s_2 while M, M' are two arbitrary corresponding points). From this figure it is clear that if M tends to S , then M' tends to T_2 .

Thus, if the common point of the lines s_1, s_2 is assigned to one of these lines, then to this point there corresponds on the second line

* We recall that all the conics in the form of oval curves are projectively equivalent.

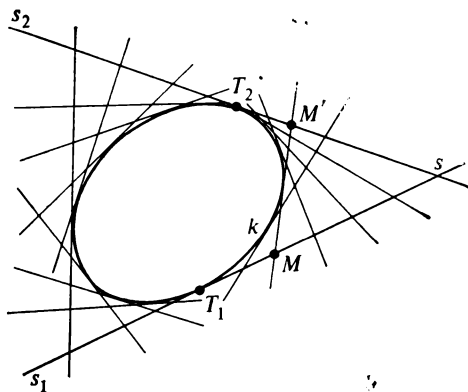


Figure 129

a point of contact with the curve k (that is, there corresponds a characteristic point of the line-conic).

This proposition is evidently the dual of the proposition established above concerning tangents to the conic having points of contact at the centres of projective pencils generating this curve (the dual association of the tangents to a curve and the characteristic points of a pencil was noted at the end of § 127).

§ 142. Let us now study Steiner's converse theorems. In the preceding subsection it was shown that every conic can be defined as a geometrical figure consisting of the common points of the corresponding rays of two projective pencils, and that this curve passes through the centres of these pencils. The following theorem establishes that the centres of projective pencils generating a conic are the ordinary points of this curve.

Theorem 57. *Let k be any nondegenerate conic and P, P' two arbitrary points on this curve. If to every ray m of the pencil P there is assigned a ray $m' = f(m)$ of the pencil P' which intersects m on k , then the correspondence $m' = f(m)$ is a projectivity.*

Remark. For degenerate conics this theorem is valid only when certain conditions are imposed on the location of the points P and P' ; namely, if k is a pair of lines, then both points must lie on one of them because otherwise the correspondence $m' = f(m)$ would not be one-to-one.

Proof. We know that the conic k may be generated by means of two projective pencils; let S and S' be the centres of the projective pencils generating k (Fig. 130). On k fix a point M and let N be any variable point on the curve, let n and n' denote the rays SN and $S'N$. By the condition imposed on S and S' , the correspondence $n' = \varphi(n)$ is a projectivity. Further, denote by U the point of intersection of

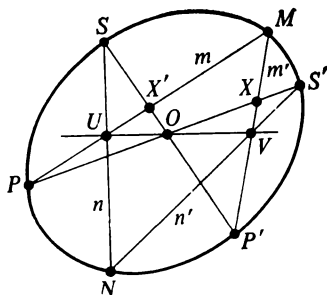


Figure 130

the line PM with the ray n and by V that of the line $P'M$ with the ray n' . Because the correspondence $n' = \varphi(n)$ is a projectivity, the correspondence assigning the point V of the line $P'M$ to the point U of PM will be a projectivity between the points of the lines PM and $P'M$. What is more, this correspondence is a perspectivity; indeed, if the point U coincides, while varying on the line PM , with M , then its corresponding point V simultaneously coincides with M and this feature ensures, in view of Theorem 53, the perspectivity of the correspondence $U \sim V$. Let us find the centre of perspective for this correspondence. To this end, note that if the point N coincides with the point P , then the point U also coincides with P and the corresponding position of V will be X , the point of intersection of the lines $P'M$ and PS' . Thus the points P and X correspond to each other; accordingly, the centre of perspective lies on the line PX , or, what is the same, on the line PS' . A similar argument shows that the centre of perspective lies on the line $P'S$. In Fig. 130 this centre is denoted by O . Thus the line UV always passes through the point O .

Now fix the point N and treat M as a variable point. We shall examine the correspondence $m' = f(m)$ between the rays originating from the points P and P' towards M and the correspondence $V = \Phi(U)$ between the points of the lines SN and $S'N$ (these lines are now fixed). The corresponding points U and V are now determined by the intersection of the lines SN and $S'N$ with the corresponding rays m, m' of the pencils P, P' . By the preceding result, the line UV always passes through O (which does not change with M). This implies that the correspondence $V = \Phi(U)$ is a perspectivity (because this is obtained by a chain of the operations of projection and section). ■

In view of the duality principle** this theorem implies

* By the hypothesis of the theorem, P and P' are arbitrary points on the curve.

** That this can be applied here follows from Theorem 52.

Theorem 58. *Let k be a nondegenerate conic and t, t' two tangents to it. If with every point M of the line t there is associated a point $M' = f(M)$ of the line t' such that the line MM' touches the curve k , then the correspondence $M' = f(M)$ is a projectivity.*

Theorems 57, 58 and 46 imply, by the way, the following two propositions that are dual to each other.

(1) *If M and M' are any points of a conic and a, b, c, d and a', b', c', d' are rays originating from these points to any points A, B, C, D of this curve (Fig. 131), then*

$$(abcd) = (a'b'c'd')$$

(2) *If m and m' are any two tangents to a conic and A, B, C, D and A', B', C', D' are points on the lines m and m' determined by the intersection with any four tangents a, b, c, d (Fig. 132), then*

$$(ABCD) = (A'B'C'D')$$

§ 143. Let us consider the construction of a conic when five of its elements are given. The results of the preceding article enable us to obtain a number of propositions described below.

(1) *Five points of the plane, no three of which are collinear, always determine a unique nondegenerate conic passing through these points.*

Indeed, let there be given five points in the plane, no three of which are collinear. Call any two of them S and S' and the remaining three A, B, C . Then from the points S and S' draw the rays a, b, c and a', b', c' toward the points A, B, C , and establish a projective correspondence between the rays of the pencils with centres S and S' such that the rays a', b', c' are assigned to the rays a, b, c . Then the geometric figure formed by the points of intersection of the corresponding rays of the pencils S and S' will be a conic which passes through the given points; there can be no other conic because the projective correspondence is uniquely determined by prescribing three pairs of corresponding elements.

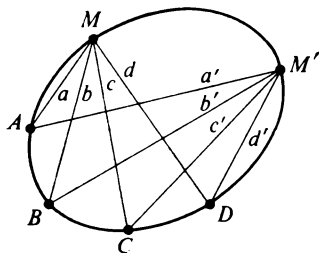


Figure 131

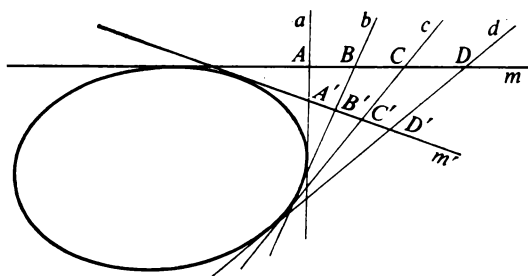


Figure 132

In § 140, we described a method for constructing corresponding pairs of rays of two projective pencils; applying this method to the present case one can obtain as many other points of a conic as one desires provided five points on this curve are known.

By the way, Fig. 130 contains a plan of the instrument for drawing a conic. In fact, suppose that for five given points S, P, N, P', S' there are established fixed bars $P'S, SN, NS'$ and $S'P$ with whom variable bars $PM, P'M, UV$ are connected by means of movable hinges at the points U, V and fixed hinges at the points P, O, P' . Then, if the bar UV rotates about the point O , the bars PM and $P'M$ connected with it rotate in such a way that their point of intersection M draws a conic that passes through the given points S, P, N, P', S' .

(2) *Four points of the plane, no three of which are collinear, and the line passing through one of them determine a unique conic that passes through the given points and touches the given line.*

Indeed, denoting the given elements as in Fig. 133, we establish a projectivity between the rays of the pencils S and S' on the basis of three pairs of corresponding rays $a, a'; b, b'; c, c'$. Then the uniquely defined, geometrical figure formed by the points of intersection of the corresponding rays of the pencils S and S' will be a conic that passes through the points A, B, S, S' and touches the line c' (see § 141).

(3) *Three noncollinear points of the plane and two lines one of which passes through one of the three given points and the other through one of the remaining two points determine a unique conic that passes through the given three points and touches the given lines.*

Indeed, denoting the given elements as shown in Fig. 134, we establish a projectivity between the rays of pencils S and S' on the basis of the three pairs of the corresponding rays $a, a'; b, b'; c, c'$. Then the geometrical figure formed by the points of intersection of the corresponding rays of the pencils S and S' will be a conic that passes through the points S, A, S' and touches the lines b and c' .

In view of the duality principle, the propositions just proved imply that a nondegenerate conic is uniquely determined, as an envelope of a line-conic, by prescribing five of its elements in one of the following combinations:

- (i) the given elements are five tangents;
- (ii) the given elements are four tangents and the point of contact on one of them;
- (iii) the given elements are three tangents and the points of contact on two of them.

§ 144. Now we shall discuss two propositions of projective geometry known as Pascal's and Brianchon's theorems.

Theorem 59 (Pascal's theorem). *Given a six-point inscribed in a conic, the points of intersection of the opposite sides are collinear (Fig. 135).*

Theorem 60 (Brianchon's theorem). *Given a six-point circumscribing a conic, the lines joining its opposite vertices are concurrent (Fig. 136).*

These theorems are evidently dual; therefore it suffices to prove only one of them.

A careful analysis of the preceding discussion reveals that Pascal's theorem is only a rewording of Steiner's theorem on the construction of a conic by means of two projective pencils and therefore has already been established in implicit form. In order to see this, first we must indicate the rule according to which one can find the pairs of opposite sides of the six-point for any given location of its vertices. To this end, we designate by 1, 2, 3, 4, 5, 6 the sides of the hexagon in the order they are successively joined; by opposite sides we shall mean the sides whose numbers differ by three, that is, the sides 1, 4; 2, 5; 3, 6. Having done this, we go back to Fig. 130. Here we have a six-point inscribed in a conic whose sides appearing in the order they are joined are SN , NS' , $S'P$, PM , MP' , $P'S$; these sides will be respectively designated by 1, 2, 3, 4, 5, 6. In Fig. 130, the point of

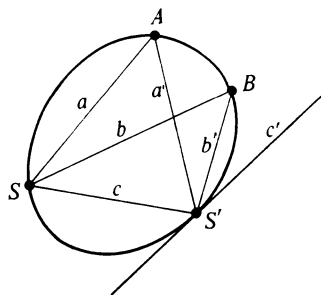


Figure 133

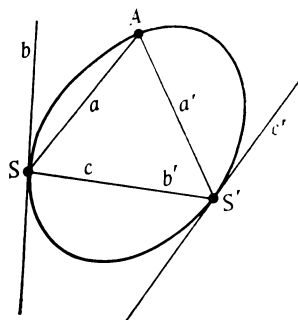


Figure 134

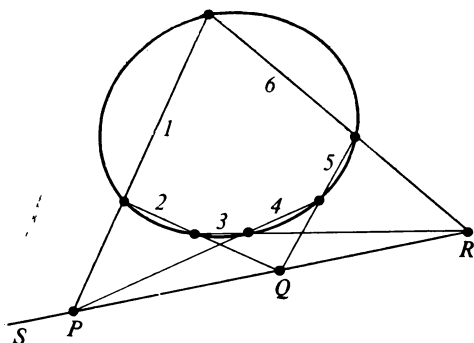


Figure 135

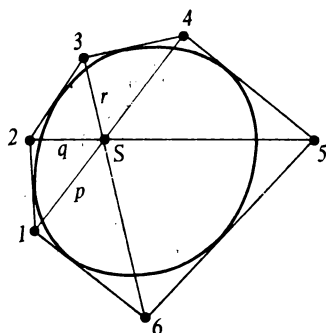


Figure 136

intersection of sides 1, 4 is U , that of sides 2, 5 is V and that of sides 3, 6 is O . It was shown earlier that the three points U, O, V are collinear and that the points S, P, N, P', S', M on the curve are totally arbitrary; thus Pascal's theorem was established there also.

As already noted, Brianchon's theorem follows from Pascal's theorem in view of the duality principle.

§ 145. We now study the limiting cases of Pascal's and Brianchon's theorems. Suppose that the points determining a certain side of the inscribed six-point coincide (for example, the points determining side 3 in Fig. 135). Then this side becomes a tangent and a figure drawn in Fig. 137 is obtained. We therefore have the theorem:

A tangent drawn to a conic at one of the vertices of the inscribed five-point intersects the side opposite to this vertex at a point which lies on the line passing through points of intersection of the remaining pairs of nonadjacent sides of the given five-point.

The dual of this proposition for the limiting case of Brianchon's theorem is obtained if we assume that two adjacent sides of the inscribed six-point coincide and their common vertex becomes the point of contact (Fig. 138). Accordingly, we have the following theorem:

The line joining the point of contact of one of the sides of the circumscribed five-point with the opposite vertex passes through the common point of the lines joining the remaining two pairs of nonadjacent vertices of the given five-point.

Other limiting cases of Pascal's theorem for an inscribed four-point and an inscribed three-point and the limiting cases of Brianchon's theorem for a circumscribed four-point and a circumscribed three-point are demonstrated in Figs. 139, 140, 141 and 142 without any explanation.

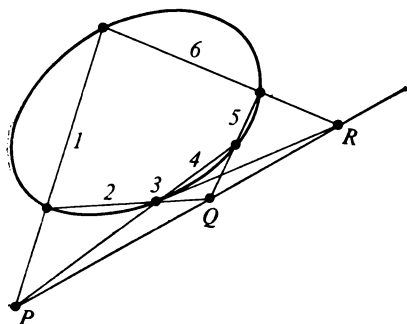


Figure 137

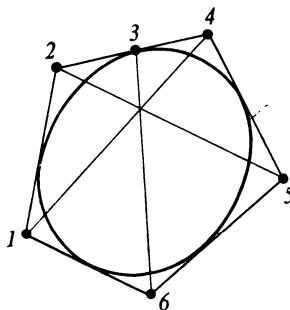


Figure 138

§ 146. Let us see how one can draw the tangent at a given point of a conic and obtain the point of contact of the given tangent.

Problem. Given five points on a conic, draw a tangent at one of them.

This problem is solved by using Pascal's theorem for an inscribed five-point. Suppose that the segments joining the given points are designated by numbers 1, 2, 4, 5, 6 as in Fig. 137 and the point where the tangent is to be drawn by 3. First find the points P , Q and then the point R and then join R with 3; this will give the desired tangent.

Problem. Given five tangents to a conic, obtain the point of contact of one of them.

This problem is solved using Brianchon's theorem for a circumscribed five-point. Let the points of intersection of the given tangents be denoted by the numbers 1, 2, 4, 5, 6 as in Fig. 138. Then by joining the points 1, 4 and 2, 5 by lines we obtain the point of intersec-

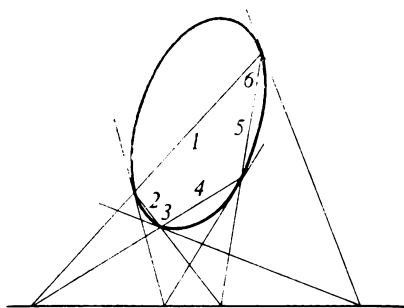


Figure 139

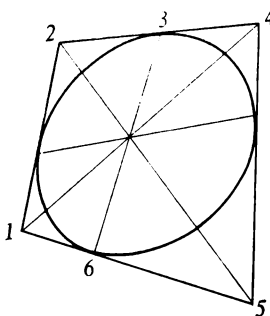


Figure 140

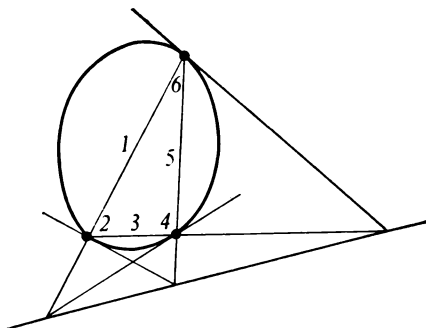


Figure 141

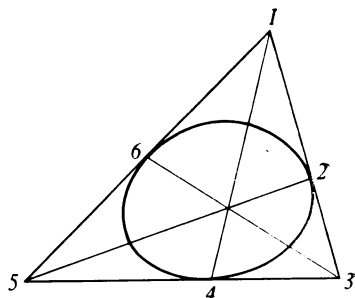


Figure 142

tion of these lines, and the point of intersection of the lines joining this point with 6 and 2 with 4 is the desired point of contact.

§ 147. In the present section we have examined in detail the projectivity between the elements of one-dimensional manifolds of the first degree. For many problems of projective geometry it proves useful to extend the notion of a projectivity for the set of elements of one-dimensional manifolds of the second degree. We shall now describe how such a generalization can be effected by taking oval curves as a concrete case of the manifold of the second degree.

Two pairs of points A, B and C, D of a conic k will be said to be harmonically conjugate if they are projected from a point M of k by two pairs of harmonically conjugate rays. In view of the first of the two propositions stated at the end of § 142, we may assert that the harmonic conjugateness of the two pairs A, B and C, D of the conic does not depend on the point M chosen and thus is determined solely by the location of points A, B, C, D .

A one-to-one correspondence between the points of two distinct or coincident conics k_1 and k_2 is called a projectivity if under this correspondence harmonically conjugate pairs of points on k_2 are assigned to harmonically conjugate pairs of points on k_1 .

We find that this definition is completely analogous to the definition of a projectivity between the points of lines. The above projectivity will also be called a *projective mapping* of the conic k_1 into the conic k_2 . When k_1 and k_2 coincide the conic is said to be projectively mapped into itself. We will deal with this case below.

Suppose that a projectivity is defined of a conic k into itself. Take a point A on k and denote its image by A' . We establish a correspondence between the rays of pencils with centres A and A' by associating a ray m of the pencil A with a ray m' of the pencil A' such that the point of intersection of m with k is the image of the point of intersection of m' with k . It is easily seen that such a correspondence is a projectivity. Indeed, if m', n' and p', q' are harmonically conjugate pairs of rays of pencil A' , then, according to the definition of harmonic conjugateness on a conic, the pairs of points M, N and P, Q , where the rays m', n', p', q' intersect the curve, are also harmonically conjugate. Under a projectivity of a conic into itself the pairs of points M, N and P, Q are mapped into harmonically conjugate pairs M', N' and P', Q' which are projected from the point A by harmonically conjugate pairs of rays m, n and p, q . But these very rays of pencil A are associated with the rays m', n', p', q' of pencil A' . Thus under the established correspondence a set of harmonic elements of pencil A is assigned to the set of harmonic elements of pencil A' , and this is what characterizes a projectivity.

One can say more about the established correspondence between the rays of pencils A and A' : this correspondence is not only a projectivity but a perspectivity too. This is a consequence of the fact that the ray AA' of pencil A is assigned to the ray $A'A$ of pencil A' (see Theorem 54).

Therefore the corresponding rays of the pencils A and A' intersect on one line, the axis of perspective of these pencils. This gives a very simple method for a graphical realization of the projectivity of a conic into itself, which can be applied to the case where this mapping is determined by three pairs of corresponding points (this method also contains the proof of the fact that three pairs of corresponding points determine a projectivity). Consider three pairs of projectively corresponding points on a conic: A, A' ; M_1, M'_1 ; and M_2, M'_2 (Fig. 143). First we shall obtain the axis of perspective of the pencils A and A' , for which purpose we obtain the point of intersection

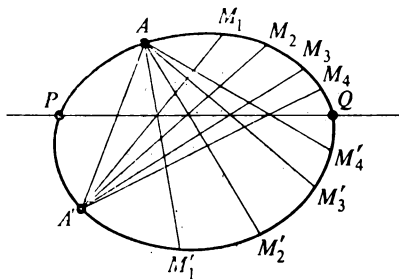


Figure 143

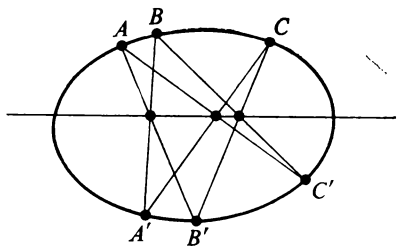


Figure 144

of the lines AM'_1 and $A'M_1$ and that of the lines AM'_2 and $A'M_2$; the line that joins these points of intersection is the axis of perspective. Having found the axis of perspective, we can find the corresponding point M' for any point M of the conic by projecting M from A' to the axis of perspective and then projecting the resulting point on the axis of perspective from A to the conic. Figure 143 shows the construction of the points $M'_3, M'_1, \dots$ corresponding to the points $M_3, M_4, \dots$.

We also note that the points of intersection of the axis of perspective of pencils A and A' with the conic are fixed points of the projectivity of the conic into itself. In fact, the above construction of projectively corresponding points on the conic immediately implies that if the point M lies both on the conic and the axis of perspective, then the corresponding point coincides with this point, that is, M is mapped into itself. In Fig. 143 one can see that when the variable point of the conic approaches point Q , where the axis of perspective intersects the conic, the corresponding point also approaches Q .

When the variable point coincides with Q , this point coincides with its corresponding point. Therefore a fixed point of the mapping is also known as a *double point*.

It is clear from what has been said so far that for obtaining double points of a projectivity of a conic into itself one is led to the construction of the axis of perspective of pencils A and A' . If the axis of perspective of these pencils does not intersect the conic the projectivity has no double points.

It is a striking fact that the axis of perspective for the pencils A and A' coincides with the axis of perspective for any other pair of pencils that project corresponding points of the conic and have centres at two corresponding points. Indeed, let A, B, C be three points on the conic and A', B', C' the corresponding points under a projectivity of this conic into itself (Fig. 144). For the centres of pencils projecting the corresponding points of the conic first take the points A and A' . The axis of perspective of these pencils s_a is determined by the point of intersection of the lines $A'B$ and AB' and that of the lines $A'C$ and AC' . Then choose the points B and B'

as centres of the projecting pencils. The axis of perspective of these pencils s_b is determined by the point of intersection of the lines $B'A$ and BA' and that of the lines $B'C$ and BC' . But by Pascal's theorem the three points thus obtained are collinear and accordingly the axes s_a and s_b coincide.

The common axis of perspective for the pencils A and A' , B and B' , etc. is simply referred to as the *axis of perspective for the projectivity of a conic into itself*.

All that has been said so far enables us to state the following proposition.

If a projectivity of a conic into itself is given, then, given any two pairs of corresponding points M , N and M' , N' , the lines MN' and $M'N$ always intersect on a definite line, which is the axis of perspective for the given mapping.

§ 148. To obtain the double points of a projectivity of a line into itself. The results of the preceding article enable us to solve the following problem.

Problem. Obtain the double points of a projectivity that maps a line a into itself when three pairs of corresponding points A , A' ; B , B' ; C , C' of this mapping are given.

Solution. Using a compass, draw a circle k (Fig. 145) and take any point S on it. Next define a correspondence between the points of k by assigning to any point M^* of this circle a point $M^{*'}$ such that the pair M^* , $M^{*'}$ is projected from S into a pair of points of a line a which correspond to each other under the given projectivity. Evidently, the correspondence $M^* \sim M^{*'}$ is a projectivity (on

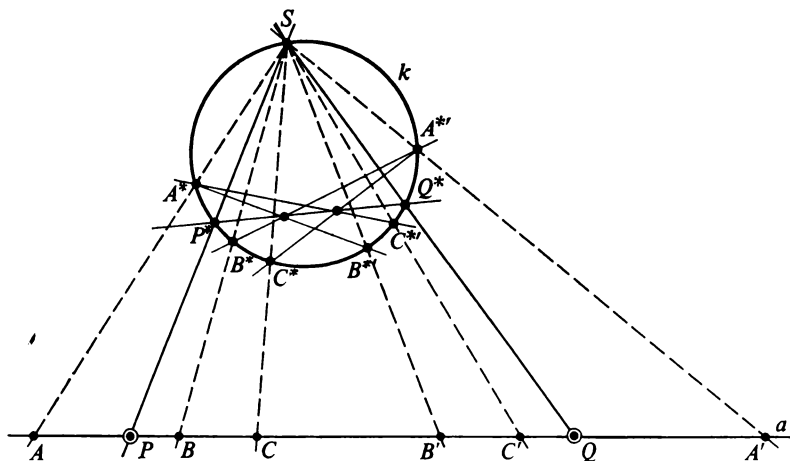


Figure 145

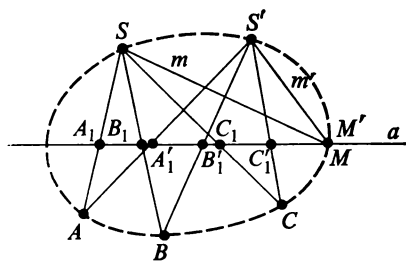


Figure 146

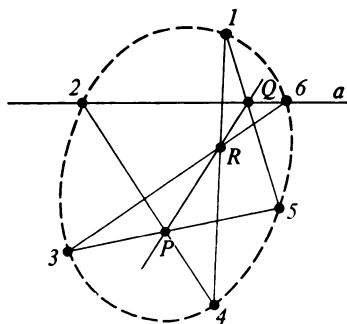


Figure 147

the circle) and its double points are projected from S into the double points of the given projectivity of line a into itself.

Thus the problem reduces to obtaining double points of the projectivity of circle k into itself. To obtain these points, project the points A, B, C, A', B', C' from S onto the circle k ; this gives the points $A^*, B^*, C^*, A'^*, B'^*, C'^*$ on the circle. Next, construct the axis of perspective for the correspondence $M^* \sim M'^*$ defined by three pairs of corresponding points A^*, A'^* ; B^*, B'^* ; C^*, C'^* and then locate the points P^*, Q^* where the axis of perspective thus constructed meets the circle. Projecting the points P^*, Q^* from centre S onto line a , we obtain the desired double points P, Q . (All the constructions have been shown in Fig. 145.)

§ 149. Let us now obtain the points of intersection of any line with a conic determined by five points.

Problem. A conic k is determined by five points S, S', A, B, C (Fig. 146). Find the point of intersection of this conic with any line a (the conic k is not assumed to have been drawn, only its points S, S', A, B, C are known).

This problem reduces to the preceding one. Indeed, the conic k can be thought of as a geometrical figure composed of points of intersection of projectively corresponding rays m, m' of the pencils with centres S, S' , provided a correspondence $m \sim m'$ is defined such that the rays $S'A, S'B, S'C$ correspond to the rays SA, SB, SC . By intersection with line a , the pairs of corresponding rays m, m' determine pairs of projectively corresponding points M, M' ; in particular, the pairs of corresponding rays $SA, S'A; SB, S'B; SC, S'C$ determine the pairs of corresponding points $A_1, A'_1; B_1, B'_1; C_1, C'_1$ on line a . The points of intersection of a and k are the points where the corresponding rays m, m' meet, and, accordingly, they are the double points of the projective correspondence $M \sim M'$. Thus to solve the problem we must obtain double points of the

projectivity of line a into itself assuming that this projectivity is determined by three pairs of points $A_1, A'_1; B_1, B'_1; C_1, C'_1$. This has been done in the preceding article.

Remark. If the line a passes through one of the given five points, the problem of finding the only unknown point of intersection with k is considerably simplified. In this case one can use Pascal's theorem. In Fig. 147, the given five points of k are designated by 1, 2, 3, 4, 5, the desired point of intersection of a with k by 6, and P, Q, R denote the points of intersection of the opposite sides of the hexagon 1, 2, 3, 4, 5, 6. Because this hexagon is inscribed in the conic k , the points P, Q, R are collinear; therefore point 6 can be obtained in the following manner: first find the points P, Q and then point R by drawing the line PQ , and, finally, point 6 by joining the points 3 and R by a line.

§ 150. Let us now consider the problem of how to draw tangents from a given point of the plane to a conic determined by five points.

Problem. Given a conic determined by five points A, B, C, D, E and given an arbitrary point P , draw tangents from P to k .

Solution. Through P and any two of the given five points, for example A and B , draw lines PA and PB (Fig. 148). Following the method just described, find points A_1 and B_1 , where the lines PA and PB meet k . Then construct a complete quadrangle ABA_1B_1 ; from its harmonic properties it follows that the diagonal p is the polar of P with respect to the conic k (for the definition of a polar, see § 131). Finally, the points M_1 and M_2 , where p intersects k are obtained, and then join these points with P by the lines t_1 and t_2 . Theorem 51 implies that t_1 and t_2 are the desired tangents.

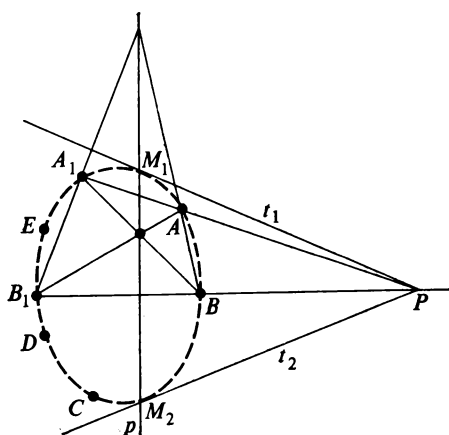


Figure 148

§ 151. We shall now present an interesting theorem, known as *Desargues' second theorem*, of projective geometry regarding pencils of conics.

By a pencil of conics we mean a set of conics which for various values of parameter λ (taking $\lambda = \infty$ also) are defined by the equation

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} + \lambda(b_{11}x^2 + 2b_{12}xy + b_{22}y^2 + 2b_{13}x + 2b_{23}y + b_{33}) = 0 \quad (*)$$

where a_{ik} , b_{ik} are constants, and x , y are current (projective) coordinates. Obviously, this pencil is the set of conics that pass through four points of intersection of the two conics

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} = 0$$

and

$$b_{11}x^2 + 2b_{12}xy + b_{22}y^2 + 2b_{13}x + 2b_{23}y + b_{33} = 0$$

The four points, called the *basis points of the pencil*, may be either real or imaginary.

Theorem 61 (Desargues). *Conics belonging to some pencil intersect a given line not passing through basis points of the pencil pairs of points that correspond to each other under an involution.*

Before proving this theorem, let us recall the definition of an involution. As defined in § 113, an involution over a line is a projectivity of this line into itself such that by double application of this mapping a point of the line is mapped into itself, that is, if $M' = f(M)$ is the image of M , then $M'' = f(M') = M$. In projective coordinates on the line the coordinates x , x' of the points M , M' , which correspond under an involution, are connected by the relation

$$x' = \frac{\alpha x + \beta}{\gamma x + \delta}$$

where $\alpha = -\delta$ (see § 113).

Now we turn to the proof of the theorem. Let a be any line mentioned in Desargues' theorem. The coordinate system on the line is assumed to be chosen in such a way that the x -axis coincides with a . Then to determine the points of intersection of the conics of the pencil and the line a it suffices to put $y = 0$ in equation (*). This yields

$$a_{11}x^2 + 2a_{13}x + a_{33} + \lambda(b_{11}x^2 + 2b_{13}x + b_{33}) = 0 \quad (**)$$

Let x , x' be the coordinates of two points M , M' , where the line a is intersected by the conic belonging to the pencil, corresponding

o a certain value of λ . By Viète's theorem, we have from (**),

$$x + x' = -\frac{2(a_{13} + \lambda b_{13})}{a_{11} + \lambda b_{11}} \quad (\alpha)$$

$$xx' = \frac{a_{33} + \lambda b_{33}}{a_{11} + \lambda b_{11}} \quad (\beta)$$

We now try to establish a relationship between x and x' . To do this, we eliminate λ from (α) and (β) and obtain

$$\begin{vmatrix} a_{11}(x+x') + 2a_{13} & b_{11}(x+x') + 2b_{13} \\ a_{11}xx' - a_{33} & b_{11}xx' - b_{33} \end{vmatrix} = 0$$

whence

$$x' = \frac{(a_{11}b_{33} - a_{33}b_{11})x + 2(a_{13}b_{33} - a_{33}b_{13})}{2(a_{13}b_{11} - a_{11}b_{13})x - (a_{11}b_{33} - a_{33}b_{11})} \quad (\gamma)$$

We find that x' is expressed in terms of x by means of a linear-fractional function; accordingly, the correspondence $M(x) \sim M'(x')$ is a projectivity. Further, by comparing formula (γ) with the general formula $x' = (\alpha x + \beta)/(\gamma x + \delta)$ we see that $\delta = -\alpha$, that is, the condition characterizing an involution is fulfilled in the given case.

We shall now show that the determinant Δ of the transformation (γ) does not vanish. But $\Delta = -(a_{11}b_{33} - a_{33}b_{11})^2 - 4(a_{13}b_{11} - a_{11}b_{13})(a_{13}b_{33} - a_{33}b_{13})$ is the resultant of the functions $a_{11}x^2 + 2a_{13}x + a_{33}$ and $b_{11}x^2 + 2b_{13}x + b_{33}$ multiplied by minus one; if $\Delta = 0$, then these functions have a common root, that is, the line a passes through a basis point of the pencil, which is ruled out by the hypothesis of the theorem. Hence $\Delta \neq 0$ and the theorem is proved. ■

The next two subsections are devoted to the applications of Desargues' second theorem.

§ 152. According to Theorem 42, § 113, an involution is determined by two distinct pairs of corresponding points. Now we shall show how, by knowing two pairs of corresponding points of the involution, to obtain many other such pairs and double points (if they exist).

Let A, A' and B, B' be two pairs of corresponding points under an involution on a line a . Through the points A and A' draw a circle k_a and through B, B' a circle k_b such that the two circles intersect at two points; call the points of intersection O_1 and O_2 (Figs. 149a and b). A system of circles passing through O_1 and O_2 is a particular case of the pencil of conics. By Desargues' second theorem, the circles belonging to this system intersect line a in pairs of points that correspond under an involution. Thus by drawing various circles through the points O_1 and O_2 and determining their points of intersection with line a we obtain various pairs of points that correspond to each other under the involution determined by the pairs A, A' and B, B' .

Drawing two circles k_m and k_n through the points O_1 and O_2 which touch the line a , we obtain the double points of the involution, namely the points of contact the circles k_m and k_n with the line a . In Fig. 149a the double points have been denoted by M and N .

If we compare Figs. 149a and b, it is easily understood that the two pairs of points A, A' and B, B' determine a hyperbolic involution (that is, one having double points) if these pairs do not divide each other, and an elliptic (that is, one having no double points) otherwise.

§ 153. Let us now determine a conic by prescribing four of its points and one tangent line.

Problem. Given four points and a tangent line to a conic, find the point of contact of the given tangent.

This problem can be looked upon as that of determining a conic by prescribing four of its points and one tangent; indeed, having located the point of contact of the given tangent, we know five points of the conic, and a conic is completely determined by five points.

Solution. Let A, B, C, D be the four given points of the desired conic and t the given tangent to it. Consider the pencil of conics with basis points A, B, C, D . By Desargues' theorem, the conics

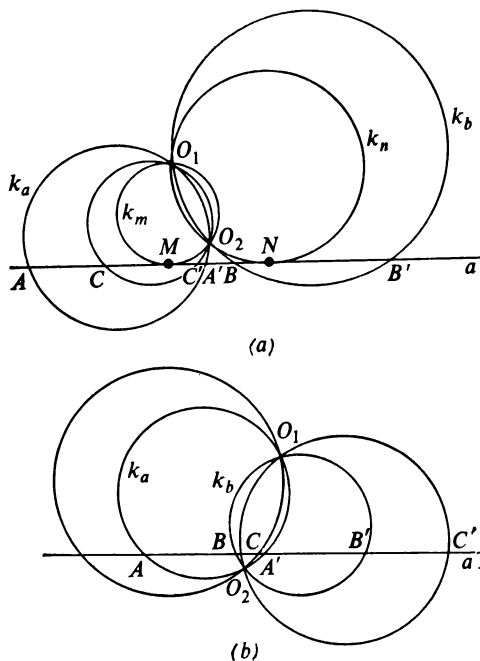


Figure 149

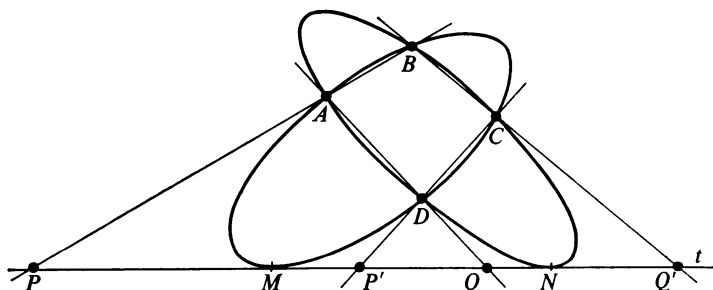


Figure 150

belonging to this pencil intersect the line t at pairs of points that correspond to each other under an involution. The desired conic is contained in the pencil mentioned and the point of contact is the double point of this involution. Thus the problem reduces to that of finding double points of the involution. In order to find these points, we have to know two pairs of corresponding points. We shall obtain them by intersecting t with any two conics of the pencil. This can be most conveniently done by taking two degenerate conics of the pencil, for example, the pair of lines AB , CD and the pair of lines AD , BC (Fig. 150).

Denote by P , P' and Q , Q' the pairs of points where these degenerate conics intersect line t . If the pairs P , P' and Q , Q' do not divide each other, then by the method of the preceding article two double points M , N of the involution are obtained. Each of these points is the point of contact with line t of a certain conic passing through the given points A , B , C , D . In this case the problem has two solutions.

If the pairs P , P' and Q , Q' divide each other, then the involution defined by them has no double points. Therefore in this case the problem has no solution (in the real plane).

§ 154. In the present section we have set out a number of concrete theorems on projective properties of conics. The source of these properties is the construction, described above, of conics by means of two projective pencils. It is natural to inquire if the same methods can be extended to the theory of curves of higher orders. In principle, this is possible. For example, a third order curve can be constructed by means of two projective pencils, one of which is a pencil of conics and the other a pencil of lines. How this can be done will now be described. Let

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 + 2a_{13}x + 2a_{23}y + a_{33} \\ + \lambda (b_{11}x^2 + 2b_{12}xy + b_{22}y^2 + 2b_{13}x + 2b_{23}y + b_{33}) = 0 \quad (*)$$

be a pencil of conics and

$$y - y_1 = \lambda' (x - x_1) \quad (**)$$

a pencil of lines. With a conic belonging to the pencil (*) that corresponds to a certain value of parameter λ we associate that line in the pencil (**) which corresponds to the value of parameter λ' given by the formula

$$\lambda' = \frac{\alpha\lambda + \beta}{\gamma\lambda + \delta} \quad (***)$$

where $\alpha, \beta, \gamma, \delta$ are constants such that $\alpha\delta - \beta\gamma \neq 0$. Such a correspondence between the elements of the pencils (*) and (**) will be called a *projectivity*.

If the parameters λ and λ' are eliminated between the equations (*), (**) and (***), we obtain an equation

$$\Phi(x, y) = 0$$

of the third degree in x, y . Hence we have the following statement:

The geometrical figure formed by the points of intersection of the corresponding curves of two projective pencils, the one of which is a pencil of conics and the other a pencil of lines, is a third order curve.

Generalizing the notion of a projectivity to the case of two pencils of conics, one can similarly define constructively curves of the fourth order, etc.

The projectivities between the pencils of the curves of the first, second and higher orders can be put in a purely geometrical language; thus one can simultaneously give constructive and pure geometrical definition of the images of higher degrees. Working along these lines, concrete properties of images of higher degrees were investigated by several authors, but this is by no means as simple and visual as the investigation of conics; and due to its special character, this falls outside the scope of the present book.

Chapter 6

Group-Theoretic Principles in Geometry. Groups of Transformations

6.1. Geometry and the theory of groups

§ 155. In many sections of the preceding chapters of the book where the equivalence (equality) of geometric figures was defined in various geometric systems (elementary geometry, projective geometry), we noted the so-called *group properties* of the set of those transformations, which were taken as the basis of the definition of equivalent images (the group properties of motions, projective transformations). In all these cases we had occasion to employ the manifestation of the group-theoretic principles in geometry that were developed by the Norwegian analyst, geometer, and group theorist Marius Sophus Lie (1842-1899) and Felix Klein.

In contemporary geometrical investigations the group-theoretic principles play an extremely important role, to which the present chapter is devoted.

§ 156. The main topic of discussion of this chapter will be groups of transformations. But before defining them, we define a group in general. A *group* is the set of objects of arbitrary nature, to be henceforth called elements and to be denoted by $a, b, c, d, \dots$, that satisfy the following axioms:

(1) *To every pair of elements in the set taken in a definite order there is assigned according to some rule an element in this set. If element c is assigned to two elements a, b , then we shall write*

$$c = ab$$

and call c the product of a and b .

(2) *Given any three elements a, b, c ,*

$$(ab)c = a(bc)$$

always (the associative law).

(3) *There is an element e such that for any element a*

$$ae = a$$

The element e is referred to as the identity of the group.

(4) *For every element a there is an element x , depending on a , such that*

$$ax = e$$

The element x is called the inverse of a and is denoted by a^{-1} .

It is a simple matter to derive the following theorems in connection with these axioms.*

(a) If $ax = e$, then $xa = e$. In view of this property of a group, the "right" and "left" inverses need not be distinguished.

(b) If e is the identity in the group, then $ea = a$ for any a . In view of this property of a group, the "right" and "left" identities need not be distinguished.

(c) If $ax = e$ and $ay = e$, then $x = y$; that is, the inverse is uniquely determined by the element a .

As a consequence of these theorems, one has the following result: given elements a, b , the element x satisfying the equation $ax = b$ is always uniquely determined and $x = a^{-1}b$; the same is true of the element y satisfying the equation $ya = b$, namely $y = ba^{-1}$. Thus, in a group an operation inverse of the group multiplication is always uniquely defined.

If the elements e and e^* satisfy the equations $ae = a$ and $ae^* = a$ for any a , then $e^* = e$, that is, a group contains only one identity.

We present one more definition. If a certain part of the elements of a group also satisfies the group axioms, then this part is called a *subgroup* of the group in question; evidently, a subgroup always contains the identity of the group and the inverse of every element of the subgroup.

In this article we have dealt with an abstract group for which the nature of its elements and that of the group multiplication do not play any role in its theory; only the requirements of Axioms 1-4 are important. From now on we shall deal only with concrete *groups of transformations*, the general definition of which is given in the next subsection.

§ 157. Let M be any set whose elements are denoted by $x, y, z, \dots$ or by $x', y', z', \dots$, etc. If to each element x in M there is assigned an element $x' = f(x)$ in the same set, then we shall say that a transformation of M into itself is defined.

If (1) to distinct elements x_1, x_2 there always correspond distinct elements $x'_1 = f(x_1), x'_2 = f(x_2)$, and (2) for every element x' in M there is an element x such that $x' = f(x)$, that is, every element in M is the image of some element of this set, then $x' = f(x)$ is known as a one-to-one transformation of M .

Let $x' = f(x)$ be a one-to-one transformation of M . If to every element y in M there is assigned that unique element y' which, according to the stated rule, is mapped into y , that is, y' is such that $f(y') = y$, then we obtain a new one-to-one transformation

* See, for instance, L. S. Pontrjagin: *Topological Groups*, Oxford University Press, London, 1946.

$y' = \varphi(y)$ of the set M . This transformation is called the *inverse* of the given transformation.

Thus every one-to-one transformation $x' = f(x)$ has a definite inverse transformation (also one-to-one). The inverse of the given transformation is usually denoted by $x' = f^{-1}(x)$.

Let $x' = f_1(x)$ and $x' = f_2(x)$ be two one-to-one transformations of M . If to every element y in M there is assigned that element y' into which y is mapped by performing the first of the two given transformations and then the second, that is, the element $y' = f_2(y^*)$, where $y^* = f_1(y)$, then we obtain a new one-to-one transformation. This transformation is called the *product* of the two given transformations (taken in a definite order), and is written $x' = f_2(f_1(x))$.

In general, the product of transformations depends on the order in which they are performed, that is, $f_2(f_1(x)) \neq f_1(f_2(x))$, in general.

The transformation $e(x) = x$, which maps all the elements into themselves, is known as the *identity transformation*. Obviously, if $x' = f(x)$ is a one-to-one transformation and $x' = f^{-1}(x)$ is its inverse, then $f(f^{-1}(x)) = x = e(x)$ and $f^{-1}(f(x)) = x = e(x)$, that is, the product of a given transformation and its inverse is an identity transformation (and the order in which the given transformation and its inverse are performed is immaterial).

Let a set M be given. We consider all possible one-to-one transformations of this set; as usual, they are denoted by $x' = a(x)$, $x' = b(x)$, $x' = f(x)$, etc. or, what is more convenient for the present situation, simply by $a, b, f, \dots$, etc. If a and b are two transformations $x' = a(x)$ and $x' = b(x)$, then their product can be represented either by $x' = a(b(x))$ or by $x' = b(a(x))$, depending on which of the transformations is performed first. In light of this, $c = ab$ will denote the product of a, b if b is performed first and $c = ba$ the product of a, b if a is performed first.

It is easy to show that the set of all one-to-one transformations of M forms a group if the product of two elements of this set, that is, of two transformations is understood in the sense described above. Indeed,

(i) With every pair of transformations a, b , taken in a definite order, there is defined a new transformation c , their product:

$$c = ab$$

(ii) If a, b, c are any three transformations, then

$$(ab)c = a(bc)$$

This equation is obvious. Indeed, if $x' = a(x)$, $x' = b(x)$ and $x' = c(x)$ are given transformations, then both $(ab)c$ and $a(bc)$ denote the transformation $x' = a(b(c(x)))$. In this way, the product of transformation always obeys the associative law.

(iii) There is a transformation e (namely, the identity transformation $e(x) = x$) such that

$$ae = a$$

for any transformation a . Indeed, if $x' = a(x)$ is the transformation denoted by a and $e(x) = x$ is the identity transformation, then ae is the transformation defined by $x' = a(e(x)) = a(x)$, which is the same as a .

(iv) For any transformation a there is a transformation f such that

$$af = e$$

Namely, for f one can take the inverse of the given transformation a , that is, $f = a^{-1}$.

We see that the set of all one-to-one transformations of M satisfies the group axioms 1-4. Accordingly, this set is a group. The identity for this group is the identity transformation. Besides the group of all transformations of M , any other definite set of transformations satisfying the group axioms is also called the group of transformations of that set.

For a definite set of transformations of M to be a group it is enough that the following two conditions be satisfied:

(a) If a, b are transformations belonging to the given set of transformations, their product ab also belongs to this set.

(b) If a is any transformation belonging to the given set, its inverse a^{-1} also belongs to this set.

Indeed, as noted above, the associative law for the product of transformations always holds; moreover, if the given set contains the inverse a^{-1} together with every transformation a and together with any two transformations their product, then the transformation $e = aa^{-1}$ (the identity of the group of transformations) is contained in this set. Therefore, if the set of transformations satisfies the above two conditions, it satisfies all the group axioms and is hence a group.

§ 158. Let us now study the geometry of a given group. Let a set M of arbitrary elements and a group G of its transformations be given. We shall call M a *space* and its elements *points* and every point set of this set a *figure*. A figure A will be called *equivalent*, or equal, to a figure B if there is a transformation in G that maps A into B .

The two conditions characterizing a group of transformations, which were formulated at the end of § 157, immediately imply that

(1) If a figure A is equivalent to a figure B , then B is equivalent to A .

Indeed, if A is equivalent to B , then there is a transformation g in G that maps A into B . In that case, the inverse transformation g^{-1} maps B into A . But by the second of the two conditions men-

tioned above, g^{-1} belongs to G . Thus there is a transformation in G that maps B into A , and therefore B is equivalent to A .

(2) *If two figures A and B are equivalent to a third figure C , then the figures A and B are equivalent to each other.*

Indeed, if A is equivalent to C , then there is a transformation g in G mapping A into C , and if B is equivalent to C , there is a transformation h in G mapping B into C . Then the transformation $h^{-1}g$ maps C into B and therefore the product $h^{-1}g$ maps A into B , and the equivalence of A and B follows.

We find that the conditions defining the group of transformations are necessary to ensure the basic properties of the equivalence of figures (reflexivity and transitivity) without which the word "equivalent" would be meaningless.

Following Felix Klein, by *geometrical properties* we mean those properties of figures in space M and those quantities connected with figures that remain invariant under any transformation belonging to a given group G and are therefore the same for all the equivalent figures. The system of propositions regarding properties of figures and quantities which remain invariant under all the transformations belonging to the group G will be called the *geometry of Group G* .

The ideas propounded by Klein for treating various geometries as theories of invariants under corresponding groups enabled him to unfold a deep connection between geometries, discoveries and researches up to the 1880s. Klein advanced his ideas in his lectures "A Comparative Review of Recent Researches in Geometry", which is known by the name *Erlanger Programm*.

Concrete applications of Klein's group-theoretic methods are given in the next section.

6.2. The projective group and its basic subgroups

§ 159. In the preceding subsection the concept of geometry of a group was defined. The definition is very general because no conditions on space M and group G were imposed. It is clear that the geometry of a given group G will be meaningful only if this group G and space M on which the group is defined are given sufficiently concrete form. Henceforth we shall confine ourselves to the study of the *geometry of the projective group*.

The investigation which we shall make will show all the various geometries studied in previous chapters in a new light and as parts of a definite system. In order not to complicate the presentation with involved algebraic computations, we shall carry out our investigation for the two-dimensional case. And because we shall be employing an exclusively analytical method, the results obtained can be easily

extended to higher dimensions. To accomplish this, one has only to replace the necessary equations with equations of the same structure but having more number of variables. Such a modification can be easily done by the reader himself.

§ 160. Let us now study the projective group. Consider the projective plane, that is, the set of points determined by the triples of homogeneous coordinates (x_1, x_2, x_3) . The one-to-one mapping of the plane into itself under which to a point $M(x_1, x_2, x_3)$ there corresponds the point $M'(x'_1, x'_2, x'_3)$ with coordinates

$$\left. \begin{aligned} \rho' x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho' x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho' x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3 \end{aligned} \right\} \quad (1)$$

where c_{ik} are real constants such that

$$\Delta = \begin{vmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{vmatrix} \neq 0$$

and $\rho' \neq 0$, is a projective mapping, or, alternatively, a projective transformation (projectivity) of the projective plane.

It was shown in § 106 that the set of projective transformations possesses the group property. Namely, by Theorem 23a, the inverse of a projectivity is also a projectivity, while, by Theorem 23b, the product of two projectivities is a projectivity. This, in view of § 157, enables us to assert that the set of projectivities is a group.

Observe that the group properties of the set of projectivities can be easily established by a pure analytical method (in § 106 they were established geometrically). Indeed, let

$$\rho'_1 x'_i = \sum_{\alpha=1}^3 c_{i\alpha}^{(1)} x_\alpha \quad (i=1, 2, 3) \quad (2)$$

$$\rho'_2 x'_i = \sum_{\alpha=1}^3 c_{i\alpha}^{(2)} x_\alpha \quad (i=1, 2, 3) \quad (3)$$

be the analytic representations of two projectivities $M' = f_1(M)$ and $M'' = f_2(M')$. Consider any point $M(x_1, x_2, x_3)$. The first transformation maps this point into a point $M'(x'_1, x'_2, x'_3)$ and the second maps $M'(x'_1, x'_2, x'_3)$ into a point $M''(x''_1, x''_2, x''_3)$. From (2) and (3) we have

$$\rho_2 x''_i = \sum_{\alpha=1}^3 c_{i\alpha}^{(2)} x'_\alpha = \sum_{\alpha=1}^3 c_{i\alpha}^{(2)} \left(\frac{1}{\rho'_1} \sum_{\beta=1}^3 c_{\alpha\beta}^{(1)} x_\beta \right) = \frac{1}{\rho'_1} \sum_{\beta=1}^3 \left(\sum_{\alpha=1}^3 c_{i\alpha}^{(2)} c_{\alpha\beta}^{(1)} \right) x_\beta$$

If we set $\rho'_1 \rho'_2 = \rho''$ and $\sum_{\alpha=1}^3 c_{i\alpha}^{(2)} c_{\alpha\beta}^{(1)} = c_{i\beta}$, then the last equation may be written as

$$\rho'' x_i = \sum_{\beta=1}^3 c_{i\beta} x_{\beta} \quad (i = 1, 2, 3) \quad (4)$$

which shows that the transformation $M'' = f_2(M') = f_2(f_1(M))$, that is, the product of two given transformations, has the same analytical structure as equation (1). Furthermore, let $C^{(1)}$, $C^{(2)}$ and C denote the matrices composed of the quantities $c_{ih}^{(1)}$, $c_{ih}^{(2)}$ and c_{ih} , respectively. In view of the relation $\sum_{\alpha=1}^3 c_{i\alpha}^{(2)} c_{\alpha h}^{(1)} = c_{ih}$, the matrix C is the product of the matrices $C^{(1)}$ and $C^{(2)}$:

$$C = C^{(2)} C^{(1)} \quad (5)$$

Thus the product of two projectivities (2) and (3) is a linear transformation (4) whose matrix equals the product of the matrices of transformations (2) and (3).

Denoting by $\Delta^{(1)}$, $\Delta^{(2)}$ and Δ the determinants of the matrices $C^{(1)}$, $C^{(2)}$ and C , we see that formula (5) implies that

$$\Delta = \Delta^{(1)} \Delta^{(2)} \quad (6)$$

Hence, if $\Delta^{(1)} \neq 0$ and $\Delta^{(2)} \neq 0$, $\Delta \neq 0$. This shows that the product of two projectivities is a linear transformation with a nonvanishing determinant, that is, a projectivity.

In order to show that the inverse of a projectivity is also a projectivity, it suffices to note that from equations (1) the quantities x_1, x_2, x_3 can be expressed as linear combinations of x'_1, x'_2, x'_3 provided $\Delta \neq 0$. Furthermore, the linear transformation obtained by reciprocating (1) has a matrix which is obviously the inverse of the matrix of (1) and corresponding determinant $\Delta' = 1/\Delta$. Hence $\Delta' \neq 0$. Since the inverse of a projectivity is a linear transformation with a nonvanishing determinant, it is a projectivity. Thus by simple algebraic computations we find that the set of projectivities is a group because this set satisfies the two conditions that, according to § 157, characterize the group of transformations.

The group of projective transformations is known as the *projective group*. Each separate transformation in this group is determined by assigning definite values to the c_{ih} in (1). But since the relationships in (1) are homogeneous, to define transformation (1) it is sufficient to specify eight ratios of c_{ih} . These eight ratios are known as the *parameters* of the projective group.

In the case where each transformation entering into a group, no matter which one, is determined by assigning numerical values for n independent parameters, the group is called an *n-member group*, and the projective group (on a plane) is an eight member group.

§ 161. Projective geometry is a discipline concerned with the study of those properties of figures and those quantities connected with the figures that are invariants relative to any projective transformation. We can therefore define projective geometry as the geometry of projective groups.

In the study of projective geometry particular attention is given to the *invariants of the projective group*, since in this branch of geometry it is these invariants which represent the geometric quantities.

We shall give the name invariant for n arbitrary points relative to the projective group to a scalar function $F(M_1, M_2, \dots, M_n)$ which is not identically a constant but which has equal values on such systems of n points that convert to one another through projective transformations.*

We note that a projective group does not have invariants for three or less points, this can easily be proved indirectly. Let us assume that there exists an invariant for three points, $F(M_1, M_2, M_3)$. We take any three points M_1^0, M_2^0, M_3^0 on a plane and denote the value of the function $F(M_1^0, M_2^0, M_3^0)$ by c . Let M_1, M_2, M_3 be any three points. We know that no matter where the points M_1, M_2, M_3 and M_1^0, M_2^0, M_3^0 lie there is a projective transformation that transforms the points M_1, M_2, M_3 into M_1^0, M_2^0, M_3^0 . Hence $F(M_1, M_2, M_3) = c = \text{constant}$, contrary to the definition of an invariant. In exactly the same manner, it may be shown that the projective group has no invariants of four arbitrary points. But in a projective group there is an invariant of four collinear points: this is the cross ratio (see § 115).

When $n \geq 5$, there are invariants for any system of n points. It is a striking fact that all of them may be expressed in terms of the cross ratios. We shall clarify the situation by taking the most simple case $n = 5$.

Let M_1, M_2, M_3, M_4, M_5 be five arbitrary points in the plane. By means of the construction shown in Fig. 151, we can determine two quadruples of points M_1, Q, M_2, P and M_5, R, M_4, P lying on one line each, on the basis of the five given points M_1, M_2, M_3, M_4, M_5 . Evidently, the cross ratio $(M_1QM_2P) = f_1$ and $(M_5RM_4P) = f_2$ are functions of the points M_1, M_2, M_3, M_4, M_5 , which will be expressed by writing

$$f_1 = f_1(M_1, M_2, M_3, M_4, M_5)$$

and

$$f_2 = f_2(M_1, M_2, M_3, M_4, M_5)$$

* In order to avoid having to account for all the various possible special cases, let us assume, that speaking of arbitrary points whose number is greater than two always implies a group of points such that no three lie on a single straight line.

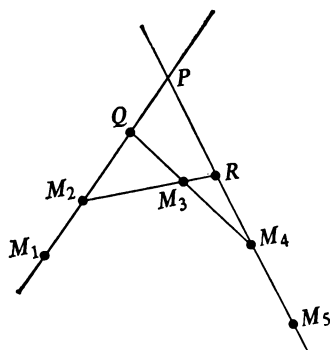


Figure 151

These functions are invariants of the projective group. Indeed, let $M'_1, M'_2, M'_3, M'_4, M'_5$ be five new points and P', Q', R' three points determined by M'_i in the same manner as P, Q, R are determined by the points M_i . If a projectivity maps M_1, M_2, M_3, M_4, M_5 into $M'_1, M'_2, M'_3, M'_4, M'_5$, then it maps P, Q, R into P', Q', R' and hence $(M_1QM_2P) = (M'_1Q'M'_2P')$ and $(M_5RM_4P) = (M'_5R'M'_4P')$. Thus as soon as the system $M_1, \dots, M_5$ is projectively equivalent to the system $M'_1, \dots, M'_5$, we have $f_1(M_1, \dots, M_5) = f_1(M'_1, \dots, M'_5)$ and $f_2(M_1, \dots, M_5) = f_2(M'_1, \dots, M'_5)$. And this means that f_1 and f_2 are indeed projective invariants.

What is more, the converse can also be proved; that is, if $f_1(M_1, \dots, M_5) = f_1(M'_1, \dots, M'_5)$ and $f_2(M_1, \dots, M_5) = f_2(M'_1, \dots, M'_5)$, then the system of five points $M_1, \dots, M_5$ is projectively equivalent to the system $M'_1, \dots, M'_5$. Indeed, let $M_1, \dots, M_5$ and $M'_1, \dots, M'_5$ be two systems of points satisfying $f_1(M_1, \dots, M_5) = f_1(M'_1, \dots, M'_5)$ and $f_2(M_1, \dots, M_5) = f_2(M'_1, \dots, M'_5)$. We can find a projectivity $M' = \varphi(M)$ that maps M_1, M_2, M_4, M_5 into the four points M'_1, M'_2, M'_4, M'_5 and the point P into P' . In view of the condition given, $f_1(M_1, \dots, M_5) = f_1(M'_1, \dots, M'_5)$, that is $(M_1QM_2P) = (M'_1Q'M'_2P')$, and therefore $M' = \varphi(M)$ must map Q into Q' . Similarly, it follows from $f_2(M_1, \dots, M_5) = f_2(M'_1, \dots, M'_5)$ that $M' = \varphi(M)$ maps R into R' . But then $M' = \varphi(M)$ evidently maps M_3 into M'_3 , establishing the equivalence of the systems $M_1, \dots, M_5$ and $M'_1, \dots, M'_5$.

Suppose that $F(M_1, M_2, M_3, M_4, M_5)$ is any projective invariant of five points. Taking an arbitrary system of points M_1, M_2, M_3, M_4, M_5 , we shall deform this system so that the quantities $f_1(M_1, M_2, M_3, M_4, M_5)$ and $f_2(M_1, M_2, M_3, M_4, M_5)$ remain unchanged. According to the preceding result, all the systems result-

ing from such a deformation are equivalent to the original system and therefore $F(M_1, M_2, M_3, M_4, M_5)$ does not change its value. Thus if f_1 and f_2 have definite numerical values, then F also has a definite numerical value and therefore F is a function of f_1 and f_2 , that is, F has the form $F = \Phi(f_1, f_2)$.

By similar arguments it can be proved that any projective invariant $F(M_1, \dots, M_n)$ with $n \geq 5$ is expressed in terms of cross ratios. On this account the cross ratio is referred to as the *basic invariant* of the projective group.

§ 162. Let a group of transformations G of an arbitrary space M be given. The transformations belonging to G which map a point set U in M into itself are called *automorphic transformations* or, in short, *automorphisms* with respect to the set U ; an automorphism may displace the points of U , but only into points of U .

The set of all transformations belonging to a group G which are automorphisms with respect to U is a group.

Let us show this.

(1) If each of the two transformations in G maps U into itself, then their product is in G and has the same property, that is, the product of two automorphisms is an automorphism.

(2) If a transformation in G maps U into itself, then the inverse transformation belongs to G and has the same property, that is, the inverse of an automorphism is an automorphism.

In view of § 157, these properties characterize the group property of the set of automorphisms.

§ 163. In the projective plane take any line and call it the line at infinity and denote it by ∞ . The set of projectivities which are automorphisms with respect to the line ∞ is, by the preceding result, a subgroup of the projective group. We shall call it the *affine group* and every transformation belonging to this group the *affine transformation*.

Obviously, affine transformations map the finite points of the projective plane (that is, points not lying on ∞) into finite points. Therefore the affine transformations are also one-to-one transformations of the set of finite points of the projective plane, that is, one-to-one transformations of the projective plane cut along the line ∞ .

The projective plane without the line ∞ will be called the *affine plane*.*

We shall now obtain analytic representation for affine transformations. To this end, we introduce homogeneous projective coordinates (x_1, x_2, x_3) on the projective plane so that in these coordi-

* In its topology the affine plane does not differ from the Euclidean plane.

nates the equation of line ∞ is $x_3 = 0$. Let the formulas

$$\left. \begin{aligned} \rho' x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho' x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho' x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3 \end{aligned} \right\} \quad (*)$$

define a certain projectivity. This transformation will be an affine transformation if $x_3 = 0$ implies $x'_3 = 0$ for any x_1, x_2 (that is, if the line $x_3 = 0$ is mapped into itself). And for this it is both necessary and sufficient that the coefficients c_{31} and c_{32} vanish. Thus we have the following representations for affine transformations:

$$\left. \begin{aligned} \rho' x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho' x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho' x'_3 &= c_{33}x_3 \end{aligned} \right\} \quad (**)$$

Since for every finite point $x_3 \neq 0$, the entire affine plane may be arithmetized by the nonhomogeneous coordinates $x_1/x_3 = x, x_2/x_3 = y$. Therefore the use of homogeneous coordinates in the study of the affine group is meaningless. We shall obtain analytic representation in nonhomogeneous coordinates for the affine group if the first and second equations in (**) are divided by the third and then the notation $x_1/x_3 = x, x_2/x_3 = y, x'_1/x'_3 = x', x'_2/x'_3 = y'$ be introduced. Setting further $c_{i1}/c_{33} = a_i, c_{i2}/c_{33} = b_i, c_{i3}/c_{33} = c_i$, we can write

$$\left. \begin{aligned} x' &= a_1x + b_1y + c_1 \\ y' &= a_2x + b_2y + c_2 \end{aligned} \right\} \quad (A)$$

Every transformation of form (A) is affine provided $\Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \neq 0$, otherwise the transformation would not be one-to-one. Since (A) contains six parameters, the affine group is a six member group.

§ 164. We shall now study the invariants of the affine group. The geometry of the affine group is known as the *affine geometry*.

The affine geometry which studies the properties of figures and quantities associated with these figures that are invariant under the affine group differs considerably from projective geometry. For example, whereas any two lines intersect in projective geometry (in the projective plane), there exist parallel lines in affine geometry (in the affine plane). Namely, the lines in the projective plane intersecting at a point of the line ∞ turn into parallel lines in the affine plane when the projective plane is cut along the line ∞ (the common point is removed in the cut). Obviously, the Euclidean parallel postulate holds in affine geometry: through any point outside a given line there passes one and only one line parallel to the given line. Note also that, as on a Euclidean line, linear ordering of points holds on an affine line (see § 94).

We shall now examine the question of invariants of an affine group, that is, the geometric quantities from the point of view of affine geometry.

First of all note that all the projective invariants are also the affine invariants. Indeed, if a function is invariant under all projectivities, it remains invariant under all affine transformations because affine transformations constitute a part of projectivities. In contrast, there are affine invariants that are not projective invariants.

The basic affine invariant is the simple ratio of three collinear points. The simple ratio of three points $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, $M_3(x_3, y_3)$, denoted by $(M_1M_2M_3)$ may be defined by any one of the following formulas*:

$$(M_1M_2M_3) = \frac{x_2 - x_1}{x_3 - x_2}, \quad (M_1M_2M_3) = \frac{y_2 - y_1}{y_3 - y_2}$$

To prove the invariance of $(M_1M_2M_3)$ is a simple matter. In fact, let M'_1, M'_2, M'_3 be three points into which a transformation of the form (A) maps the points M_1, M_2, M_3 . If x'_i, y'_i denote the coordinates of points M'_i , formulas (A) yield

$$\begin{aligned} x'_2 - x'_1 &= a_1(x_2 - x_1) + b_1(y_2 - y_1) \\ &= (M_1M_2M_3)[a_1(x_3 - x_2) + b_1(y_3 - y_2)] \\ x'_3 - x'_2 &= a_1(x_3 - x_2) + b_1(y_3 - y_2) \end{aligned}$$

whence

$$(M'_1M'_2M'_3) = \frac{x'_2 - x'_1}{x'_3 - x'_2} = (M_1M_2M_3)$$

which means that the function $(M_1M_2M_3)$ is invariant under any affine transformation. There are no affine invariants of points situated arbitrarily (not on one line). This is due to the fact that an affine transformation maps any three noncollinear points into three noncollinear points (it was proved in § 161 that the projective group does not have invariants of four arbitrary points; the proof of the fact that the affine group does not have invariants of three arbitrary points is analogous). When $n \geq 4$, there are affine invariants of an arbitrary system of n points. It is a striking fact that all of them may be expressed in terms of simple ratios (this may be proved by arguments analogous to the ones used in § 161). On this account, the simple ratio of three collinear points is referred to as the *basic invariant of the affine group*.

Remark. The affine plane and hence affine geometry may be defined in complete independence from projective geometry by means of a suitable system of axioms. Namely, by an affine plane we mean

* That these formulas determine the same quantity can be easily shown by using the equation of a line $y = kx + l$.

the set of objects of two kinds, points and lines that satisfy axioms belonging to five groups, of which

the first group defining the incidence relation between the objects contains the first three axioms of Group I from the system of the axioms of Euclidean geometry (§ 12) (that is, the two-dimensional axioms of Group I);

the second group defining the order of points on a line coincides with the second group of axioms of Euclidean geometry (since the linear order of points holds on an affine line, the affine axioms of "betweenness" must coincide with the axioms of "betweenness" of Euclidean geometry);

the third group contains Dedekind's axiom of continuity;

the fourth group contains Euclid's axiom of parallels;

the fifth group contains Desargues' theorem (the formulation of which must be modified in view of the fact that parallel lines exist).*

For defining an affine space all the axioms of three-dimensional Euclidean geometry, except the axioms of congruence, are to be adopted.

It was shown in the appropriate place that the axioms on which elementary geometry is based constitute a complete system. In the same way, one can prove that the system of axioms of affine geometry is complete. At the same time, the affine axiom system is a part of the Hilbert system of axioms. At first this might appear to be paradoxical but its explanation is not difficult.

The thing is that the completeness of the affine axioms, which means that any of its realizations is isomorphic to some definite realization (arithmetical, for example) does not hinder the addition of new axioms of congruence to the affine axioms because together with these axioms a new relationship between the geometric objects (namely, the congruence relationship) is also introduced. In this connection, see the definition of the completeness of a system of axioms stated in § 75.

§ 165. The affine transformation

$$\left. \begin{aligned} x' &= a_1x + b_1y + c_1 \\ y' &= a_2x + b_2y + c_2 \end{aligned} \right\} \quad (*)$$

is called *unimodular* if

$$\Delta = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = \pm 1$$

It can be easily demonstrated that the unimodular affine transformations constitute a group. Let us show this.

(1) The product of two unimodular affine transformations is a unimodular affine transformation. To prove this, note that if the

* See D. Hilbert: *Foundations of Geometry*.

transformation

$$x'' = a_1x + b_1y + c_1$$

$$y'' = a_2x + b_2y + c_2$$

is a product of the transformations

$$x' = a_1^{(1)}x + b_1^{(1)}y + c_1^{(1)}$$

$$y' = a_2^{(1)}x + b_2^{(1)}y + c_2^{(1)}$$

and

$$x'' = a_1^{(2)}x' + b_1^{(2)}y' + c_1^{(2)}$$

$$y'' = a_2^{(2)}x' + b_2^{(2)}y' + c_2^{(2)}$$

then the matrices of these transformations are connected by the relation

$$\begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} = \begin{pmatrix} a_1^{(2)} & b_1^{(2)} \\ a_2^{(2)} & b_2^{(2)} \end{pmatrix} \cdot \begin{pmatrix} a_1^{(1)} & b_1^{(1)} \\ a_2^{(1)} & b_2^{(1)} \end{pmatrix}$$

Hence the determinants of these matrices satisfy $\Delta = \Delta^{(1)}\Delta^{(2)}$. Accordingly $\Delta = \pm 1$ if $\Delta^{(1)} = \pm 1$ and $\Delta^{(2)} = \pm 1$.

(2) The inverse of a unimodular affine transformation is a unimodular affine transformation.

To prove this, it is enough to note that inverse affine transformations have mutually inverse matrices and, accordingly, reciprocal determinants, that is, if Δ_1 is the determinant of the given transformation and Δ_2 that of the inverse transformation, then $\Delta_2 = 1/\Delta_1$. Hence $\Delta_2 = \pm 1$ if $\Delta_1 = \pm 1$.

We see that the set of unimodular affine transformations satisfy the two conditions which, according to § 157, define the group character of the set of transformations. Thus the unimodular affine transformations indeed constitute a group. This group as well as the geometry based on this will be called *unimodular affine*.

The unimodular affine group is a five member group because in this case the six parameters present in formulas (*) are connected by one equation $a_1b_2 - a_2b_1 = \pm 1$, and therefore only five of the parameters are independent.

Obviously, all the objects of the general affine geometry are also the objects of the unimodular affine geometry, but the latter contains objects not belonging to the former since the class of invariants of the unimodular affine group is wider than that of the general affine group.

We shall now show that the unimodular affine group includes the invariant of three points located arbitrarily. Suppose that any three points $M_1(x_1, y_1)$, $M_2(x_2, y_2)$, $M_3(x_3, y_3)$ in the affine plane are mapped into three points $M'_1(x'_1, y'_1)$, $M'_2(x'_2, y'_2)$, $M'_3(x'_3, y'_3)$ by a unimodular affine transformation. Then it is a simple matter to

verify that

$$\begin{vmatrix} 1 & 1 & 1 \\ x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} \cdot \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = \pm \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

This shows that the absolute value of the determinant $\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$ is invariant of three points M_1, M_2, M_3 .

In the unimodular affine geometry, we can associate the invariant

$$S = \text{absolute value of } \frac{1}{2} \begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

with every triangle $M_1M_2M_3$. The quantity S is called the *area of the triangle* $M_1M_2M_3$. Obviously, in unimodular affine geometry one can also define the area of a polygon and the area of a curvilinear figure. Namely, the area of a polygon may be defined as the sum of the areas of triangles constituting this polygon and the area of a curvilinear figure as the limit of the sequence of the areas of polygons approximating the given figure. Thus the area of figures are objects of the unimodular affine geometry.

§ 166. The affine transformation

$$\left. \begin{aligned} x' &= a_1x + b_1y + c_1 \\ y' &= a_2x + b_2y + c_2 \end{aligned} \right\} \quad (1)$$

is called *orthogonal* if its matrix

$$A = \begin{pmatrix} a_1 & b_1 \\ a_2 & b_2 \end{pmatrix} \quad (2)$$

satisfies the condition

$$AA' = I \quad (3)$$

where the prime denotes the transpose of A and I is the identity matrix, that is,

$$A' = \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

We shall show that the set of orthogonal transformations constitutes a group.

(1) *The product of two orthogonal transformations is an orthogonal transformation.*

Proof. Let two orthogonal transformations with matrices A_1 and A_2 be given; their product is an affine transformation with the matrix $A = A_2A_1$. Applying the rule for multiplication of matri-

ces, we have the identity

$$AA' = (A_2A_1)(A_2A_1)' = (A_2A_1)(A_1'A_2') = A_2(A_1A_1')A_2'$$

This, in view of the relations $A_1A_1' = I$ and $A_2A_2' = I$, implies that

$$AA' = A_2IA_2' = A_2A_2' = I$$

as required.

(2) *The inverse of an orthogonal transformation is an orthogonal transformation.*

Proof. Let A be the matrix of an orthogonal transformation and $B = A^{-1}$ the matrix of the inverse of the given transformation. The orthogonality condition $AA' = I$ implies that $A' = A^{-1}$ and, accordingly, $B = A'$. Hence

$$BB' = A'(A')' = A'A = A^{-1}A = I$$

as required.

Thus the set of orthogonal transformations is a group; this group will be called the *orthogonal group*.

Condition (3) implies that the determinant of matrix A equals ± 1 , from which we conclude that the orthogonal group is a subgroup of the unimodular group.

The orthogonality condition (3) written in the matrix form is equivalent to three scalar relations

$$\left. \begin{aligned} a_1^2 + b_1^2 &= 1 \\ a_2^2 + b_2^2 &= 1 \\ a_1a_2 + b_1b_2 &= 0 \end{aligned} \right\} \quad (4)$$

Because the orthogonal group is formed from the affine group by means of three equations connecting the six parameters a_i, b_i, c_i , this is a three member group.

The orthogonality condition (4) may be put in another form. Indeed, as shown above (while proving the orthogonality of the inverse of the given transformation), the matrix A of the orthogonal transformation satisfies

$$A'A = I$$

Therefore

$$\left. \begin{aligned} a_1^2 + a_2^2 &= 1 \\ b_1^2 + b_2^2 &= 1 \\ a_1b_1 + a_2b_2 &= 0 \end{aligned} \right\} \quad (4')$$

The systems (4) and (4') are equivalent.

Unlike all the groups discussed so far, the orthogonal group has an invariant of two points. Such is, for example,

$$\rho(M_1, M_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

a function of two points $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$. The proof of the invariance of this function requires only simple computations. Namely, let $M'_1(x'_1, y'_1)$ and $M'_2(x'_2, y'_2)$ be two points obtained from the points $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ by an orthogonal transformation. Then

$$\begin{aligned}\rho(M'_1, M'_2) &= \sqrt{(x'_2 - x'_1)^2 + (y'_2 - y'_1)^2} \\ &= \sqrt{[a_1(x_2 - x_1) + b_1(y_2 - y_1)]^2 + [a_2(x_2 - x_1) + b_2(y_2 - y_1)]^2} \\ &= \sqrt{(a_1^2 + a_2^2)(x_2 - x_1)^2 + 2(a_1b_1 + a_2b_2)(x_2 - x_1)(y_2 - y_1) + (b_1^2 + b_2^2)(y_2 - y_1)^2}\end{aligned}$$

whence, noting (4'), we find that

$$\rho(M'_1, M'_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = \rho(M_1, M_2)$$

In the geometry of the orthogonal group the quantity $\rho(M_1, M_2)$ is known as the *distance* between the points M_1 and M_2 . The distance is the basic invariant of this geometry because all the remaining invariants may be expressed in terms of distance.

It hardly requires any explanation that the geometry of the orthogonal group is elementary (Euclidean) geometry.

§ 167. Let us now compare the various geometries. We have examined the projective group and its main subgroups: affine, unimodular affine, and orthogonal. To these groups there correspond projective, affine, unimodular affine and elementary (Euclidean geometries).

Of the groups mentioned the widest happens to be the group lying at the root of projective geometry (the projective group) and the narrowest the group lying at the root of elementary geometry (the orthogonal group). But at the same time, of these geometries projective geometry is the poorest regarding the class of objects and elementary geometry is the richest. In elementary geometry one may examine both the affine objects (the simple ratio of three collinear points, parallelism, etc.) and projective objects (the cross ratio of four points, etc.). In contrast, affine properties of figures are not studied in projective geometry, and metric properties, that is, properties depending on the measurement of segments, are not taken into account in affine geometry.

In general, the wider the group lying at the root of a geometry the narrower the class of geometric objects. This is understandable because the larger the number of transformations contained in a group the smaller the number of relationships and functions remaining invariant under all its transformations. But in this connection it should be pointed out that the properties of figures and quantities connected with these figures which remain invariant under a group

are more "stable" than the properties of figures and quantities which are invariant under any of its subgroup because the former sustain a variety of transformations without undergoing any change.

6.3. Geometries of Lobachevsky, Riemann and Euclid in the projective setting

§ 168. In this section we will show that Euclidean, Lobachevskian and Riemannian geometries are geometries of certain groups of projective automorphisms.

Consider a nondegenerate conic k in the projective plane. We shall examine the group of projective automorphisms with respect to k , that is, the group of projectivities that map k into itself (that the set of arbitrary automorphisms constitutes a group was proved in § 162).

We shall prove the following two important theorems.

Theorem A. *If k is an oval curve and A, A' any two points in the interior of k , then there are just two automorphisms with respect to k such that A is mapped into A' and any given direction at A into a given direction at A' .*

Theorem B. *If k is the null conic and A, A' are any two points in the projective plane, then there are just two automorphisms with respect to k such that A is mapped into A' and any given direction at A into a given direction at A' .**

Proof of Theorem A. Let a and a' be lines passing through A and A' in given directions (Fig. 152). Denote by C and C' the poles of these lines with respect to k , by B the point where the polar of A meets a and by B' the point where the polar of A' meets a' . The three-point ABC is self-polar with respect to k , that is, all its sides are polars of the opposite vertices. The same is true of the trivertex $A'B'C'$.

Taking the three-point ABC as the coordinate three-point with $A(0, 0, 1)$, $B(1, 0, 0)$, $C(0, 1, 0)$, we introduce on the plane a homogeneous projective coordinate system x_1, x_2, x_3 . In these coordinates the conic k has the equation (see § 134)

$$a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 = 0$$

By choosing the unit point $E(1, 1, 1)$ suitably, this equation can be reduced to the form

$$x_1^2 + x_2^2 - x_3^2 = 0 \quad (1)$$

Note that the terms containing the first two coordinates must appear with the same signs in the equation because $A(0, 0, 1)$ lies in the

* For the definition of the null conic and an oval curve see § 134; in Theorem B the projective plane should be understood to be augmented by imaginary elements, otherwise the concept of the null conic is meaningless.

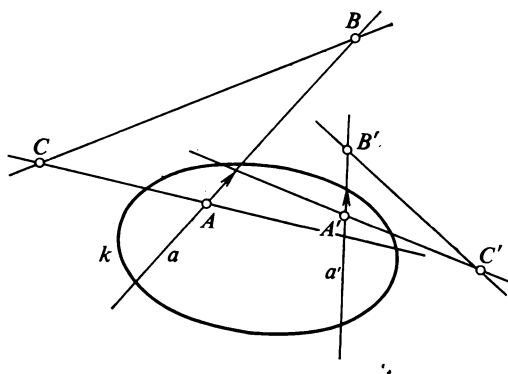


Figure 152

interior of k (for this point the left side of equation (1) is negative; see § 134).

Next, we introduce a new homogeneous projective coordinate system x'_1, x'_2, x'_3 by taking for the coordinate three-point the three-point $A'B'C'$ such that its vertices have the following coordinates: $A'(0, 0, 1)$, $B'(1, 0, 0)$, $C'(0, 1, 0)$. If the unit point $E'(1, 1, 1)$ is suitably chosen, the equation of k in new coordinates will be

$$x_1'^2 + x_2'^2 - x_3'^2 = 0 \quad (2)$$

(In this equation only the terms containing the first two coordinates must appear with the same signs because the point $A'(0, 0, 1)$ lies in the interior of k .)

Suppose that there is an automorphism with respect to k that maps A into A' , line a into line a' and the given direction into a given direction (this latter means that the points lying on a in a cyclic order are mapped into points lying on a' in a cyclic order). Because k is mapped into itself, the pole of a with respect to k must be mapped into the pole of a' with respect to k and the polar of A into the polar of A' ; in other words, the points A, B, C are mapped into the points A', B', C' , respectively. In that case the automorphism φ can be represented by the formulas

$$\rho'x'_1 = c_{11}x_1, \quad \rho'x'_2 = c_{22}x_2, \quad \rho'x'_3 = c_{33}x_3 \quad (3)$$

where x_1, x_2, x_3 are the old coordinates of the preimage and x'_1, x'_2, x'_3 are the new coordinates of the image point. Using formulas (3), equation (2) becomes

$$c_{11}^2x_1^2 + c_{22}^2x_2^2 - c_{33}^2x_3^2 = 0 \quad (4)$$

This is the equation of the preimage of k in old coordinates. Since φ is an automorphism with respect to k , equations (1) and (4) must

be equivalent, and this implies that

$$|c_{11}| = |c_{22}| = |c_{33}|$$

The quantity ρ' in (3) can be chosen arbitrarily; accordingly, one may assume without loss of generality that the moduli of c_{11} , c_{22} , c_{33} are unity and that c_{11} is positive. To make the following formulas definite, suppose that the given direction at A goes toward those points of AB for which $x_1/x_3 > 0$ and the given direction at A' goes toward those points of $A'B'$ for which $x'_1/x'_3 > 0$. Then necessarily $c_{33} > 0$, and we are left with just two alternatives: either (i) $c_{11} = 1$, $c_{22} = 1$, $c_{33} = 1$ or (ii) $c_{11} = 1$, $c_{22} = -1$, $c_{33} = 1$. Thus there can be only two automorphisms with respect to k satisfying the hypothesis of the theorem:

$$(i) \quad \rho'x'_1 = x_1, \quad \rho'x'_2 = x_2, \quad \rho'x'_3 = x_3 \quad (5)$$

$$(ii) \quad \rho'x'_1 = x_1, \quad \rho'x'_2 = -x_2, \quad \rho'x'_3 = x_3 \quad (6)$$

But it is obvious that each of these projectivities is indeed an automorphism with respect to k and that each of them maps A into A' , line a into line a' and the given direction on a into a given direction on a' . Thus Theorem A is proved. ■

Proof of Theorem B. It can be proved by repeating almost word by word the preceding proof after replacing equation $x_1^2 + x_2^2 - x_3^2 = 0$ with the equation $x_1^2 + x_2^2 + x_3^2 = 0$ (in applying the previous arguments to Theorem B the statements regarding the terms appearing in the equation with the same sign have to be dropped because these statements become meaningless as all the terms of the equation $x_1^2 + x_2^2 + x_3^2 = 0$ have the same sign). ■

Remark. The proof of Theorem A shows that every automorphism with respect to an oval curve k maps the interior points of this curve into interior points because, if $x_1^2 + x_2^2 - x_3^2 < 0$, $x_1'^2 + x_2'^2 - x_3'^2 < 0$, in view of formulas (5) and (6).

Theorems A and B can also be stated as follows:

(1) *Given any two linear elements in the interior of an oval curve k , there are two and only two automorphisms with respect to k that superpose the first linear element on the second.*

(2) *Given any two linear elements in the projective plane, there are two and only two automorphisms with respect to the given null conic k that superpose the first linear element on the second.*

Analogous properties are held by motions (including reflections) on the Euclidean plane. To emphasize this analogy, the automorphisms with respect to a nondegenerate conic k will be called *projective motions*. The conic k which is mapped into itself by the given projective motion will be called the *absolute conic* of this motion.

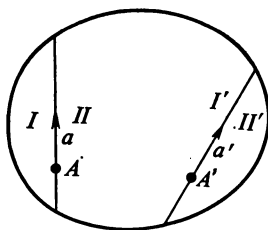


Figure 153

Motions having oval curves for the absolute conic will be termed *hyperbolic* and those having null conic for the absolute conic will be termed *elliptic*.

Figure 153 shows an oval curve and in its interior two linear elements at the points A and A' ; the lines along which these linear elements are directed are denoted by a, a' . Each of the lines a and a' divides the interior of k into two portions denoted by I, II and I', II' . From the arguments employed in proving Theorem A it follows that of the two automorphisms with respect to the absolute conic k that superpose the first given element on the second, one maps I into I' and II into II' , while the other maps I into II' and II into I' . If the point A' coincides with A and the linear element at A' with the linear element at A , then the automorphisms superposing the first linear element on the second turn into automorphisms leaving the linear element at A intact. One of these automorphisms is the identity transformation, whereas the second maps I into II and II into I . This second automorphism is analogous to the Euclidean reflection in a line.

§ 169. The geometry of the group of hyperbolic motions having a common absolute conic will be called *hyperbolic* and that of the group of elliptic motions with a common absolute conic *elliptic*.

In any of these geometries two figures are treated as equal or congruent if one of the figures can be mapped into the other by means of an automorphism with respect to the absolute conic defining the particular geometry (that is, by means of a projective motion). In both hyperbolic and elliptic geometries there exist invariants of two points. For instance, an invariant of any two points P, Q is the cross ratio $(PQUV)$, where U, V are the points of intersection of the line PQ with the absolute conic; so is any function of this cross ratio. Of special interest is the invariant of the form $c \ln (PQUV)$, where c is a constant. We shall show that this invariant possesses properties analogous to the properties characterizing the length in elementary geometry. The hyperbolic and elliptic cases will be treated separately.

We begin by examining the properties of the invariant $c \ln (PQUV)$ in hyperbolic geometry.

Let k be an oval curve serving as the absolute conic for hyperbolic geometry, and P, Q any two points in the interior of k . Since P, Q lie in the interior of k , the points U, V , where the line PQ intersects k , are real; moreover, the pair P, Q does not divide the pair U, V . With the location of points P, Q, U, V as stated, the quantity $(PQUV)$ is positive, and hence $\ln (PQUV)$ is a real number. Furthermore, if the direction of the segment PQ is opposite to that of the segment UV , $(PQUV) > 1$ and $\ln (PQUV) > 0$, whereas if the directions of the segments PQ and UV are the same, then $(PQUV) < 1$ and $\ln (PQUV) < 0$.

Assume that the first case holds. Take an arbitrary point R on the segment PQ . A straightforward calculation shows that

$$(PQUV) = (PRUV) \cdot (RQUV)$$

Taking the logarithm of both sides, we have

$$\ln (PQUV) = \ln (PRUV) + \ln (RQUV) \quad (*)$$

For the given location of points, $(PQUV) > 1$, $(PRUV) > 1$ and $(RQUV) > 1$; accordingly, all the terms in $(*)$ are positive.

If the segments PQ and UV have the same direction, then all the terms in $(*)$ are negative. In both cases, from $(*)$ we obtain

$$|\ln (PQUV)| = |\ln (PRUV)| + |\ln (RQUV)|$$

Thus, if the positive number

$$\rho (PQ) = |c \ln (PQUV)|$$

is associated with any segment PQ lying in the interior of the absolute conic k , then

(1) with congruent segments equal numbers are associated because $\rho (PQ)$ is invariant under automorphisms with respect to the absolute conic k ,

(2) the quantities associated with the segment PQ and the parts PR and RQ of this segment satisfy

$$\rho (PQ) = \rho (PR) + \rho (RQ)$$

Exactly the same properties characterize the length of a segment in elementary geometry. To underline this analogy, the positive number $\rho (PQ)$ will be called the length of the segment PQ in hyperbolic geometry with the absolute conic k .

Besides the positive number $\rho (PQ)$, with any segment PQ one may associate the relative number

$$s (PQ) = c \ln (PQUV)$$

which either coincides with the length $\rho (PQ)$ of the segment PQ or differs from it in sign provided the constant c is real.

We shall now examine the invariant $c \ln (PQUV)$ in elliptic geometry.

The absolute conic, denoted by k , for elliptic geometry is the null conic; this conic is defined in projective coordinates by an equation with real coefficients but is composed exclusively of imaginary points. Given any real points P, Q in the projective plane, the points U, V , where the line PQ meets the absolute conic, are imaginary, and the coordinates of U are the complex conjugates of those of V . It can be easily shown that then the cross ratio $(PQUV)$ is a complex number with unit modulus. In fact, if a nonhomogeneous projective coordinate system is introduced on the line PQ and if p, q, u, v denote the coordinates of P, Q, U, V , then $u = \alpha + \beta i$ and $v = \alpha - \beta i$, and

$$(PQUV) = \frac{u-p}{q-u} \div \frac{v-p}{q-v} = \frac{[(\alpha - p) + \beta i][(q - \alpha) + \beta i]}{[(\alpha - p) - \beta i][(q - \alpha) - \beta i]}$$

This shows that the cross ratio $(PQUV)$ is the quotient of two complex numbers; hence $|(PQUV)| = 1$.

Like any complex number with unit modulus, the cross ratio $(PQUV)$ may be expressed in the form

$$(PQUV) = e^{i\varphi}$$

where φ is a real quantity defined to within $\pm 2\pi k$ ($k = 1, 2, \dots$). This means that $\ln (PQUV) = i\varphi$ is a pure imaginary and many-valued quantity.

Therefore, if a pure imaginary constant c is chosen, with every segment PQ a many-valued quantity

$$s(PQ) = c \ln (PQUV) \quad (**)$$

will be associated. In order to associate one definite value with a given segment PQ , take a variable real point X on the projective line containing the segment PQ . Setting $(PXUV) = e^{i\theta}$, for $X \equiv P$ we have

$$(PPUV) = 1 \quad \text{and} \quad \theta = \theta_0 = \dots = -4\pi, -2\pi, 0, +2\pi, +4\pi, \dots$$

If X occupies any position inside the segment PQ , the equation $(PXUV) = e^{i\theta}$ determines a countable set of corresponding values for θ . When X approaches P , while remaining inside the segment PQ , each of these values approaches a definite value θ_0 . The value of θ which approaches $\theta_0 = 0$ will be denoted by $\tilde{\theta}$ and called the principal value. Furthermore, the principal value of $\ln (PQUV)$ will mean the limit to which $\tilde{\theta}$ approaches when X , while remaining inside PQ , approaches Q .

Now with every segment PQ we can associate a completely defined real number

$$s(PQ) = c \ln(PQUV) \quad (**)$$

where c is an imaginary constant and $\ln(PQUV)$ is the principal value of the natural logarithm of $(PQUV)$. Then it is evident that

(i) equal numbers are associated with congruent segments because $s(PQ)$ is invariant under automorphisms with respect to the absolute conic k ,

(ii) the numbers with the same sign associated with segment PQ and with parts PR and RQ of this segment satisfy

$$s(PQ) = s(PR) + s(RQ)$$

These properties of the invariant $s(PQ)$ provide a basis for calling the number $|s(PQ)|$ the length of the segment PQ in elliptic geometry with absolute conic k .

We note in passing that in elementary geometry the length of the entire projective line, which is equal to the length of the segment PQ with coincident endpoints, is expressed by the number $2\pi|c|$.

Having defined the length of a segment in hyperbolic and elliptic geometries, we can define the distance between two points in these geometries.

In hyperbolic geometry, the field of which is the interior of the absolute conic, the distance between two points is defined as the length of the unique segment joining these points.

In elliptic geometry, the field of which is the entire real projective plane, the distance between two points is defined as the length of the smallest of the two segments determined by these points.*

In both hyperbolic and elliptic geometries, the distance $\rho(X, Y)$ between any two points X, Y has the following properties:

- (a) $\rho(X, X) = 0$
- (b) $\rho(X, Y) = \rho(Y, X) > 0$ if $X \not\equiv Y$
- (c) $\rho(X, Y) + \rho(Y, Z) \geq \rho(X, Z)$

In other words, the quantity $\rho(X, Y)$ has the same basic properties as has distance in the Euclidean space.

The proofs of properties (a)-(c) are omitted (only the last of these needs to be proved; the first two are evident).

The lengths of segments and distances between points, which have been defined in this subsection and which are invariants under the group of automorphisms with respect to the absolute conic k ,

* We recall that the absolute conic for the elliptic geometry is the null conic which is composed of imaginary points and which does not divide the real projective plane into any regions.

are called *projective*, *elliptic* or *hyperbolic metric* depending on the nature of the absolute conic.

Remark. Since by Theorems A and B the group of automorphisms with respect to the absolute conic k is transitive with respect to linear elements, we may introduce the process of measuring lengths in both hyperbolic and elliptic geometries. To do this, first a segment AB has to be chosen as the unit of measurement. Given any other segment PQ , there is an automorphism (in view of Theorems A and B) with respect to k which maps A into P and the direction of segment AB into that of the segment PQ . Then, if B is mapped into a point P_1 lying inside the segment PQ , a segment PP_1 is marked off which is congruent to the segment AB in the sense of the geometry of the absolute conic k . Then by marking off a segment $P_1P_2 \equiv AB$ on the segment P_1Q , a segment $P_2P_3 \equiv AB$ on the segment P_2Q , etc., we can determine how many segments congruent to the segment AB are contained in the segment PQ . In this way, the integer part of the length of PQ is obtained. Then the tenths, hundredths, etc. of the length can be obtained.

It is clear that the length resulting from this process of measuring will be expressed in terms of the quantity $c \ln(PQUV)$, where U, V are the points of intersection of the line PQ with the absolute conic k . The value of c in this formula depends on the linear unit AB chosen, namely $c = 1/\ln(ABUV)$.

§ 170. We shall now show that the hyperbolic geometry in the interior of the oval curve is Lobachevskian geometry.

To this end, take an oval curve and call it k . The elements of the hyperbolic geometry defined by the absolute conic k will be called the *hyperbolic points* and the *hyperbolic lines*. The hyperbolic points are the points of the projective plane lying in the interior of k ; the hyperbolic lines are segments of projective lines lying in the interior of k , that is, chords of k . The points on k are not included in hyperbolic objects, and therefore the segments representing hyperbolic lines are open (that is, do not contain the endpoints).

The mutual incidence relationships between hyperbolic objects satisfy the Group I axioms of Euclidean geometry. Indeed, by interpreting suitably the properties of chords of a conic we find that

(1) Through any two hyperbolic points there passes a hyperbolic line, which is what Axiom I.1 requires.

(2) Through any two hyperbolic points there passes just one hyperbolic line; this is the requirement of Axiom I.2.

(3) On every hyperbolic line there exist two hyperbolic points (even an infinite number of such points); there are three hyperbolic points not lying on one given hyperbolic line. This is the requirement of Axiom I.3.

The remaining axioms in Group I are of spatial nature and need not be taken into account in two-dimensional geometry.

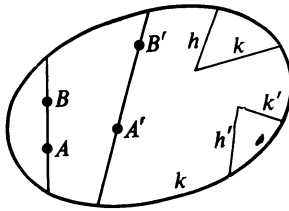


Figure 154

Next, for the hyperbolic geometry all Axioms II.1-II.3 are satisfied in the interior of k , since points on an open segment are located in a linear order. Pasch's axiom II.4 is also valid in hyperbolic geometry because this axiom is valid in the projective plane (see § 89).

Thus all the axioms of "betweenness" are true in hyperbolic geometry. Let us examine the axioms of congruence.

In Fig. 154 two hyperbolic segments AB and $A'B'$ and two hyperbolic angles (h, k) and (h', k') have been shown. In hyperbolic geometry the segment AB is regarded as congruent to the segment $A'B'$ if there is an automorphism on k that maps AB into $A'B'$; $\angle(h, k)$ is regarded as congruent to $\angle(h', k')$ if there exists an automorphism mapping hyperbolic halflines h, k into hyperbolic halflines h', k' . Theorem A of § 168 implies that on every hyperbolic line on a given side of a given point a segment can be marked off which will be congruent to an arbitrarily given segment, and that to every halfline an angle can be adjoined on a given side that will be congruent to an arbitrarily given angle.

Thus Theorem A implies the validity of Axioms III.1 and III.4 for hyperbolic geometry. By virtue of the group property of automorphisms, two segments which are congruent to a third are congruent among themselves; this shows that Axiom III.2 is also valid.

It is easy to show that all the remaining axioms of congruence are true in hyperbolic geometry (such an analysis will not be presented here).

Finally, Dedekind's continuum principle holds in hyperbolic geometry because this principle holds on every projective line. This fact and Theorem 41 (§ 23) imply that Archimedes' and Cantor's propositions are true in hyperbolic geometry.

Thus all the axioms of Groups I-IV hold for hyperbolic geometry in the interior of the absolute conic k . But then, as we know, either the Euclidean parallel axiom or the Lobachevskian parallel axiom must hold. Obviously, the latter holds. Indeed, through any point A in the interior of k there pass infinitely many chords not intersecting the given chord (Fig. 155), and this means that in hyperbolic geometry through every point there pass infinitely many lines not intersecting the given hyperbolic line.

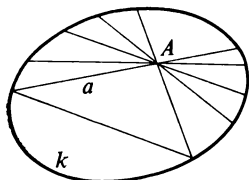


Figure 155

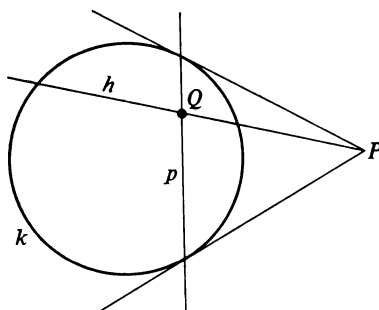


Figure 156

Summing up the above discussion, we have the following proposition: the hyperbolic geometry in the interior of the absolute conic, which is an oval curve in the given case, is the non-Euclidean geometry of Lobachevsky.

§ 171. It is interesting to see how the concrete facts of Lobachevskian geometry look when they are interpreted in the interior of the absolute conic k . We note some of them.

For example, a hyperbolic line h is perpendicular to the hyperbolic line p if, in the projective plane, h passes through the pole of p with respect to the absolute conic k .

In fact, let two hyperbolic lines h and p intersect at a point Q ; assume further that when the line h is extended beyond the interior of k , it passes through the pole P of p (Fig. 156). We perform a harmonic mapping of the projective plane into itself by choosing P as the centre of this mapping and the polar p of P as the axis. From the definitions of the polar and of the harmonic mapping (see § 131 and the Remark at the end of § 106) it follows that under this mapping the segments in the interior of k , which are divided by p , are mapped into each other. Therefore this mapping with respect to p is an automorphism which can be regarded as a reflection in p from the point of view of hyperbolic geometry.

Furthermore, it is evident that the parts of line h which are divided by Q are mapped into each other, while the line p remains fixed. Accordingly, the adjacent angles determined by the lines h and p are congruent to each other in the sense of the hyperbolic geometry of the absolute conic k , thereby meaning that h is perpendicular to k .

We note in passing that the principle of mutuality in the theory of polars (which reads thus: if one line contains the pole of another line, then the second contains the pole of the first) means the mutual perpendicularity of two lines in hyperbolic geometry (if one line is perpendicular to the other, then the second is perpendicular to the first).

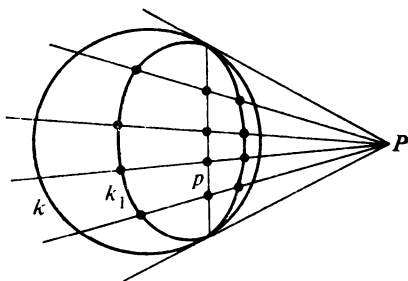


Figure 157

We devote some space to the interpretation of equidistants and oricycles in non-Euclidean geometry (see §§ 36-40).

Let k_1 be an oval curve lying in the interior of the absolute conic k and touching the absolute conic at its points of intersection with the line p (Fig. 157). Evidently, under a hyperbolic reflection in any line passing through P , which is the pole of p with respect to the absolute conic, the conic k_1 is mapped into itself. Accordingly, all the chords of k_1 directed toward P are hyperbolically congruent segments; what is more, the line p is perpendicular to these chords and divides them into two halves. Therefore, in the sense of hyperbolic geometry, the conic k_1 is an equidistant with axis p . If both points of contact of k_1 with the absolute conic coincide, then in the limit the conic k_1 turns into a oricycle. The proof of this fact is being omitted.

Several other examples of hyperbolic interpretation of non-Euclidean facts can be found in R. Baldus: *Nichteuklidische Geometrie. Hyperbolische Geometrie der Ebene*.

§ 172. We shall now show that the elliptic geometry is Riemannian geometry (see §§ 63-68).

We shall assume that the projective plane on which the elliptic geometry is defined is the plane at infinity of the Euclidean space E augmented by the elements at infinity. Suppose that a rectangular Cartesian coordinate system x, y, z with the origin at O has been introduced in the Euclidean space E . By means of these coordinates, introduce the homogeneous coordinates x_1, x_2, x_3, x_4 (see § 102). We shall assume that, apart from the elements at infinity, imaginary elements have also been adjoined to E (see § 127).

The equation $x_4 = 0$ represents the plane at infinity. In this plane the equation $x_1^2 + x_2^2 + x_3^2 = 0$ represents the null conic. This conic is taken as the absolute conic of elliptic geometry on the plane $x_4 = 0$. Simultaneously, on this plane we establish, as in § 67, basic relationships of Riemannian geometry. We have to show that these two geometries are identical.

Comparing the incidence and order relationships in these geometries, we find that they are the same (as the incidence and order relationships in projective geometry). It remains to settle the question regarding the congruence of figures; that is, we have to show that if two figures in the plane $x_4 = 0$ are congruent in the sense of elliptic geometry, they will also be congruent in the sense of Riemannian geometry and conversely. In other words, we have to show that every transformation of points of the plane $x_4 = 0$ which is a motion in the sense of elliptic geometry will also be a motion in the sense of Riemannian geometry, and conversely.

Consider a sphere k with centre at O . With an arbitrary point M in the plane $x_4 = 0$ we associate a pair of diametrically opposite points of k which are obtained by projecting M from the centre of k . With an arbitrary figure F in the plane $x_4 = 0$ we associate on the sphere k a figure composed of pairs of antipodal points which correspond to the points of F . According to § 67, two figures in the plane $x_4 = 0$ are treated as congruent in the sense of Riemannian geometry if their corresponding figures on k are congruent. This implies that on the plane $x_4 = 0$ every motion in the sense of Riemannian geometry is a transformation of points under which the images and preimages are projections of images and preimages under a certain rotation of the sphere k about the centre or under a reflection of k in the diametral plane.

Now note that every rotation of the sphere k about the centre or reflection of this sphere in the diametral plane is defined in terms of rectangular Cartesian coordinates by formulas of the form

$$\begin{aligned}x' &= c_{11}x + c_{12}y + c_{13}z \\y' &= c_{21}x + c_{22}y + c_{23}z \\z' &= c_{31}x + c_{32}y + c_{33}z\end{aligned}\tag{1}$$

where x', y', z' are the coordinates of the image point and x, y, z those of the preimage. The coefficients in (1) satisfy

$$x'^2 + y'^2 + z'^2 = x^2 + y^2 + z^2\tag{2}$$

If x, y, z are the Cartesian coordinates of a point on the sphere k , then the projection of this point onto the plane $x_4 = 0$ has for its homogeneous coordinates the numbers x_1, x_2, x_3 , which are proportional to x, y, z (see § 102). This implies that on the plane $x_4 = 0$ every motion in the sense of Riemannian geometry is defined by formulas of the form

$$\begin{aligned}\rho x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho x'_3 &= c_{31}x_1 + c_{32}x_2 + c_{33}x_3\end{aligned}$$

where x'_1, x'_2, x'_3 are the coordinates of the image point, x_1, x_2, x_3 those of the preimage and ρ is a nonzero number. Because of (2), $x'^2_1 + x'^2_2 + x'^2_3 = 0$ as soon as $x^2_1 + x^2_2 + x^2_3 = 0$. Thus on the plane $x_4 = 0$ every motion in the sense of Riemannian geometry is a projective automorphism on the null conic $x^2_1 + x^2_2 + x^2_3 = 0$. Thus we have proved that on the plane $x_4 = 0$ every motion in the sense of Riemannian geometry will also be a motion in the sense of elliptic geometry with the absolute conic $x^2_1 + x^2_2 + x^2_3 = 0$. Let us show the converse, that is, every motion in the sense of elliptic geometry is also a motion in the sense of Riemannian geometry. Consider a motion f in the sense of elliptic geometry. In the plane $x_4 = 0$ take any point M and any directed line a through M . The motion f maps the point M into a point M' and the directed line a into a directed line a' . Let f_1 and f_2 be two motions in the sense of Riemannian geometry each of which maps M into M' and a into a' . According to the result proved earlier, f_1 and f_2 are motions in the sense of elliptic geometry. But in elliptic geometry there are just two motions that map M into M' and a into a' (see § 168, Theorem B). Accordingly, f coincides either with f_1 or with f_2 ; that is, any motion taken in the sense of elliptic geometry is also a motion in the sense of Riemannian geometry. Thus on the plane $x_4 = 0$ the set of all motions in the sense of elliptic geometry coincides with the set of all motions in the sense of Riemannian geometry, proving that the two geometries are identical.

§ 173. We shall now show that Euclidean geometry is also the geometry of a certain group of projective automorphisms.

In the projective plane take any line and denote it by ∞ . On ∞ take any two imaginary points I_1 and I_2 which have complex conjugate coordinates in any system of homogeneous projective coordinates.

For the computations below, we assume the coordinate system to be so chosen that the line ∞ has the equation $x_3 = 0$ and the coordinates of the points I_1 and I_2 are $(1, i, 0)$ and $(1, -i, 0)$.

We shall examine the set of projectivities that are automorphic with respect to the pair of points I_1 and I_2 . This set, which is a group (§ 162), will be called *Klein's group*.

We try to obtain the analytic representation for Klein's automorphisms. To this end, first note that all Klein's automorphisms are at the same time automorphisms on the line $x_3 = 0$, and therefore they can be represented analytically by the formulas

$$\begin{aligned}\rho'x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho'x'_2 &= c_{21}x_1 + c_{22}x_2 + c_{23}x_3 \\ \rho'x'_3 &= c_{33}x_3\end{aligned}\tag{*}$$

(for details see § 163). We must distinguish the following two cases:

- (1) The automorphism leaves each of the points I_1 and I_2 intact;
- (2) The automorphism maps I_1 into I_2 and I_2 into I_1 .

In the first case, substituting first

$$x_1 = 1, \quad x_2 = i, \quad x_3 = 0, \quad x'_1 = 1, \quad x'_2 = i, \quad x'_3 = 0, \quad \rho' = \rho_1$$

in equation (*) and then

$$x_1 = 1, \quad x_2 = -i, \quad x_3 = 0, \quad x'_1 = 1, \quad x'_2 = -i, \quad x'_3 = 0, \quad \rho' = \rho_2$$

we obtain

$$\rho_1 = c_{11} + ic_{12}, \quad \rho_1 i = c_{21} + ic_{22}$$

and

$$\rho_2 = c_{11} - ic_{12}, \quad -\rho_2 i = c_{21} - ic_{22}$$

whence

$$c_{21} = -c_{12}, \quad c_{22} = c_{11}$$

In the second case, substituting first

$$x_1 = 1, \quad x_2 = i, \quad x_3 = 0, \quad x'_1 = 1, \quad x'_2 = -i, \quad x'_3 = 0, \quad \rho' = \sigma_1$$

in equation (*) and then

$$x_1 = 1, \quad x_2 = -i, \quad x_3 = 0, \quad x'_1 = 1, \quad x'_2 = i, \quad x'_3 = 0, \quad \rho' = \sigma_2$$

we have

$$\sigma_1 = c_{11} + ic_{12}, \quad -\sigma_1 i = c_{21} + ic_{22}$$

and

$$\sigma_2 = c_{11} - ic_{12}, \quad \sigma_2 i = c_{21} - ic_{22}$$

whence

$$c_{21} = c_{12}, \quad c_{22} = -c_{11}$$

Thus the formulas representing Klein's automorphisms necessarily have the form

$$\left. \begin{aligned} \rho' x'_1 &= c_{11}x_1 + c_{12}x_2 + c_{13}x_3 \\ \rho' x'_2 &= \mp c_{12}x_1 \pm c_{11}x_2 + c_{23}x_3 \\ \rho' x'_3 &= c_{33}x_3 \end{aligned} \right\} \quad (**)$$

in which the upper signs in the second row correspond to the case of automorphisms of the first type and the lower signs correspond to those of the second type. It is also obvious that for any values of the parameters these formulas determine Klein's automorphisms; indeed, setting $x_1 = 1$, $x_2 = \pm i$, $x_3 = 0$ in (**), we have $x'_1 \div x'_2 \div x'_3 = 1 \div \pm i \div 0$. Thus we have found an analytic representation in homogeneous coordinates for Klein's group.

With an aim to examining Klein's group on the affine plane, which is obtained by cutting the projective plane along the line $x_3 = 0$ and for all the points of which $x_3 \neq 0$, we switch over to nonhomo-

geneous coordinates. Setting

$$\frac{x_1}{x_3} = x, \quad \frac{x_2}{x_3} = y, \quad \frac{x'_1}{x'_3} = x', \quad \frac{x'_2}{x'_3} = y'$$

we divide the first two equations in (**) by the third. This yields

$$x' = \frac{c_{11}}{c_{33}}x + \frac{c_{12}}{c_{33}}y + \frac{c_{13}}{c_{33}}, \quad y' = \frac{c_{12}}{c_{33}}x \pm \frac{c_{11}}{c_{33}}y + \frac{c_{23}}{c_{33}}$$

If the new parameters are introduced by the formulas

$$\frac{c_{11}}{c_{33}} = r \cos \varphi, \quad \frac{c_{12}}{c_{33}} = r \sin \varphi, \quad \frac{c_{13}}{c_{33}} = u, \quad \frac{c_{23}}{c_{33}} = v$$

the last equations can be put in the following form

$$\begin{aligned} x' &= r(x \cos \varphi - y \sin \varphi) + u \\ y' &= r(\pm x \sin \varphi \pm y \cos \varphi) + v \end{aligned} \quad (***)$$

These formulas show that Klein's group coincides with the set of those transformations of the Euclidean plane which are a combination of motions, reflections and dilations r times. Such transformations are referred to as the *similarity transformations*.

Thus the following fundamental proposition holds:

If similar figures of the Euclidean plane are treated as congruent, then Euclidean geometry can be looked upon as the geometry of Klein's group.

Note that the set of imaginary lines passing through I_1 or through I_2 is a degenerate line-conic. Because this is mapped into itself by all transformations of Klein's group, this line-conic is called the *absolute conic* of this group. If this terminology is employed, one may state that Euclidean geometry is the geometry of the group of automorphisms with respect to a certain line-conic.

§ 174. All the discussions here concern the Euclidean plane. On this plane we introduce the rectangular Cartesian coordinates x, y and then the homogeneous coordinates x_1, x_2, x_3 , assuming that a point with Cartesian coordinates x, y has homogeneous coordinates x_1, x_2, x_3 ($x_3 \neq 0$) if $x_1 \div x_3 = x, x_2 \div x_3 = y$. Finally, we adjoin the line at infinity $x_3 = 0$ to the Euclidean plane. The points $I_1(1, i, 0)$ and $I_2(1, -i, 0)$ are called circular, or *cyclic*, points of the augmented Euclidean plane. Such a name is justified by the fact that these points are the common points of all circles. In fact, the equation of a circle in homogeneous coordinates

$$x_1^2 + x_2^2 + 2Ax_1x_3 + 2Bx_2x_3 + Cx_3^2 = 0 \quad (*)$$

is satisfied if $x_1 = 1, x_2 = \pm i, x_3 = 0$, and, accordingly, the circle (*) passes through the points I_1 and I_2 .

The imaginary points passing through a cyclic point are called *isotropic* or *minimal*.

The equation of an isotropic line passing through I_1 has the form $x_1 + ix_2 + cx_3 = 0$ and the equation of an isotropic line passing through the point I_2 has the form $x_1 - ix_2 + cx_3 = 0$. In nonhomogeneous coordinates, isotropic lines are represented by equation of the form

$$y = ix + l$$

or

$$y = -ix + l$$

It is a striking fact that the distance between any two finite points of an isotropic line is zero. Indeed, if $X_1(x_1, y_1)$ and $X_2(x_2, y_2)$ are two finite points of an isotropic line, then

$$y_2 - y_1 = \pm i(x_2 - x_1)$$

whence

$$\rho(X_1, X_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} = (x_2 - x_1) \sqrt{1 + i^2} = 0$$

It is on account of this property that the isotropic lines are called the minimal lines. Evidently, through every real point (x_0, y_0) there pass two isotropic lines

$$y - y_0 = \pm i(x - x_0)$$

which we denote j_1 and j_2 . Let u_1 and u_2 be two real lines passing through (x_0, y_0) and having angular coefficients k_1 and k_2 . Using the formula derived in § 119, we can find the cross ratio of two pairs of lines u_1, u_2 and j_1, j_2 :

$$(u_1 u_2 j_1 j_2) = \frac{i - k_1}{k_2 - i} \div \frac{-i - k_1}{k_2 + i}$$

This quantity is the invariant of Klein's group, and a natural expectation will be that this is connected with the Euclidean magnitude of the angle between the lines u_1 and u_2 . Indeed, denoting the angle $\angle(u_1, u_2)$ by φ so that

$$\tan \varphi = \frac{k_2 - k_1}{1 + k_1 k_2}$$

and transforming the right hand side as follows, we find that

$$\begin{aligned} (u_1 u_2 j_1 j_2) &= \frac{i - k_1}{k_2 - i} \div \frac{-i - k_1}{k_2 + i} = \frac{(i - k_1)(k_2 + i)}{(k_2 - i)(-i - k_1)} \\ &= \frac{k_1 k_2 + 1 - i(k_2 - k_1)}{k_1 k_2 + 1 + i(k_2 - k_1)} = \frac{1 - i \tan \varphi}{1 + i \tan \varphi} = \frac{\cos \varphi - i \sin \varphi}{\cos \varphi + i \sin \varphi} = e^{-2i\varphi} \end{aligned}$$

Hence $-2i\varphi = \ln(u_1 u_2 j_1 j_2)$ and

$$\varphi = \frac{i}{2} \ln(u_1 u_2 j_1 j_2) \quad (**)$$

The formula (**), known as *Laguerre's formula*, represents the Euclidean angle in the form of a projective invariant. This is analogous to the formulas expressing the length of a segment in hyperbolic and elliptic geometries (see § 169). The source of this analogy lies in the duality principle (for details we refer the reader to F. Klein: *Vorlesungen über nicht-euklidische Geometrie*, Chapter 6).

From the exposition of the previous subsections the reader must be convinced that the group-theoretic methods put the main geometrical systems (of Euclid, Lobachevsky and Riemann) into one setting, and this enables us to see something common to them which otherwise appears to be diversified.

Chapter 7

Minkowski Space

7.1. Multidimensional affine space

§ 175. The main subject matter of this chapter is concerned with the Minkowski space. This space is of considerable interest as a mathematical tool for physics because it is directly connected with the equations of the special relativity theory. The Minkowski space is an affine space endowed with a special metric, that is, an affine space in which the distances between points are defined in a particular manner (hence also the congruence of figures, motion, etc.).

Of special interest from the point of view of physics is the four-dimensional Minkowski space. The study of this space being the aim, we start with the theory of multidimensional affine spaces. The exposition given is based on the concept of a linear space, and is mainly independent of the previous axiomatic approach.

§ 176. Let L be a given set, and suppose that (i) a rule is given which assigns to every pair of elements x, y in L an element of the same set L ; this new element will be called the *sum* of x and y and will be denoted by $x + y$, and (ii) a rule is given which assigns to every pair x, λ , where x is an element of L and λ a real number, an element of L ; this new element will be called the *product* of x and λ and will be denoted by λx (or $x\lambda$). The operations of addition of elements in L and their multiplication by real numbers can be defined in any manner whatsoever, but the following axioms must be satisfied:

- (1) $x + y = y + x$.
- (2) $(x + y) + z = x + (y + z)$.
- (3) *Among the elements of L there is one element θ such that $x + \theta = x$ for any x ; θ is called the null element of L .*
- (4) *For every x there is an element y such that $x + y = \theta$; y is called the inverse of x and is denoted by $-x$.*
- (5) $1 \cdot x = x$.
- (6) $\alpha(\beta x) = (\alpha\beta)x$; here and later α, β denote any real numbers.
- (7) $(\alpha + \beta)x = \alpha x + \beta x$.
- (8) $\alpha(x + y) = \alpha x + \alpha y$.

The set L over elements of which the operations of addition and multiplication by real numbers are defined in such a way that the above axioms are satisfied is called a real *linear space*; the elements

of a linear space will also be referred to as *vectors*. Since we shall not be considering any other space, we shall simply speak of a linear space thereby meaning a real linear space.

One of the most simple concrete examples of a linear space is provided by the set of geometrical vectors, where the addition of elements and their multiplication by real numbers are defined in accordance with the usual rules of elementary vector algebra.

The following theorems can be derived from Axioms 1-8 (the proofs, which can be found in any book on linear algebra, are omitted here).

- (a) A linear space contains only one null element.
- (b) Every element x has only one inverse $-x$.
- (c) $0 \cdot x = \theta$ for every x .
- (d) $\theta \cdot \alpha = \theta$ for every number α .

§ 177. If the equation

$$\alpha x + \beta y + \dots + \lambda l = \theta \quad (1)$$

where $x, y, \dots, l$ are vectors and $\alpha, \beta, \dots, \lambda$ are numbers at least one of which is not zero, then the vectors $x, y, \dots, l$ are called *linearly dependent*; if (1) implies that $\alpha = 0, \beta = 0, \dots, \lambda = 0$, then the vectors $x, y, \dots, l$ are *linearly independent*.

A linear space is called an *n-dimensional space* if it contains n linearly independent vectors but any $n + 1$ vectors are linearly dependent.

Example. Consider a set K_n whose elements (vectors) are ordered sets of n real numbers: $x = \{x_1, x_2, \dots, x_n\}$. Let us define the addition of any two vectors $x = \{x_1, x_2, \dots, x_n\}$ and $y = \{y_1, y_2, \dots, y_n\}$ and multiplication of the vector $x = \{x_1, x_2, \dots, x_n\}$ with any real number λ as follows: (1) $x + y = \{x_1 + y_1, \dots, x_n + y_n\}$ and (2) $\lambda x = \{\lambda x_1, \dots, \lambda x_n\}$. It is easy to check that then all the Axioms 1-8 are satisfied (the null vector is $\theta = \{0, 0, \dots, 0\}$); if $x = \{x_1, x_2, \dots, x_n\}$ is any vector, its inverse is $-x = \{-x_1, -x_2, \dots, -x_n\}$. Therefore K_n is a linear space under the operations defined.

The space K_n contains n linearly independent vectors; for example, $\{1, 0, 0, \dots, 0\}, \{0, 1, 0, \dots, 0\}, \dots, \{0, 0, 0, \dots, 0, 1\}$. On the other hand, any $n + 1$ vectors are linearly dependent. Indeed, consider the vectors $a_1 = \{a_{11}, a_{12}, \dots, a_{1n}\}$, $a_2 = \{a_{21}, a_{22}, \dots, a_{2n}\}$, $\dots$, $a_{n+1} = \{a_{n+11}, a_{n+12}, \dots, a_{n+1n}\}$ and form the matrix

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n+11} & a_{n+12} & \dots & a_{n+1n} \end{pmatrix}$$

By a well known theorem on the rank of a matrix, the maximum number of linearly independent rows of the matrix equals the maximum number of linearly independent columns of this matrix. The number of columns in the above matrix is n ; accordingly, the number of linearly independent columns does not exceed n , and hence the number of linearly independent rows also does not exceed n . Therefore the rows of this matrix, which are $n + 1$ in number, must be connected by a linear equation, thereby meaning that the vectors $a_1, a_2, \dots, a_{n+1}$ are linearly dependent. Thus K_n contains n linearly independent vectors but any $n + 1$ vectors are linearly dependent. Consequently, K_n is an n -dimensional linear space; this space is referred to as the n -dimensional coordinate, or *arithmetic*, space.

In an n -dimensional linear space, every set of n linearly independent vectors is called a *basis*. Let $e_1, \dots, e_n$ be a basis and x any vector. Since the total number of vectors $x, e_1, \dots, e_n$ is $n + 1$, we must have

$$\alpha x + \beta_1 e_1 + \dots + \beta_n e_n = 0 \quad (2)$$

where at least one of the numbers $\alpha, \beta_1, \dots, \beta_n$ is not zero. The number α cannot be zero, since otherwise the vectors $e_1, \dots, e_n$ would be linearly dependent. Therefore we can divide by α and express (2) in the form

$$x = \left(-\frac{\beta_1}{\alpha} \right) e_1 + \dots + \left(-\frac{\beta_n}{\alpha} \right) e_n$$

or, setting $-\beta_k/\alpha = x_k$,

$$x = x_1 e_1 + \dots + x_n e_n \quad (3)$$

Formula (3), which expresses vector x , is known as the *decomposition of x with respect to the basis $e_1, \dots, e_n$* ; the quantities $x_1, \dots, x_n$ are called the *coordinates* of x relative to the basis $e_1, \dots, e_n$. It is easy to see that the decomposition of x relative to the given basis is unique. Indeed, assuming that x has another decomposition, besides (3),

$$x = x'_1 e_1 + \dots + x'_n e_n \quad (4)$$

we find from (3) and (4) that

$$(x'_1 - x_1) e_1 + \dots + (x'_n - x_n) e_n = 0 \quad (5)$$

Because the vectors $e_1, \dots, e_n$ are linearly independent, (5) implies that $x'_1 - x_1 = 0, \dots, x'_n - x_n = 0$, or $x'_1 = x_1, \dots, x'_n = x_n$, that is, decompositions (3) and (4) are not distinct.

Multiplying (3) by λ , we have

$$\lambda x = (\lambda x_1) e_1 + \dots + (\lambda x_n) e_n$$

that is, multiplication of the given vector by a number implies the multiplication of all its coordinates by the same number.

Next, suppose that any vector y has been decomposed with respect to the basis $e_1, \dots, e_n$:

$$y = y_1 e_1 + \dots + y_n e_n \quad (6)$$

Addition of (3) and (6) yields

$$x + y = (x_1 + y_1) e_1 + \dots + (x_n + y_n) e_n$$

that is, addition of vectors implies the addition of the corresponding coordinates. Thus, if a basis is chosen in an n -dimensional linear space, the vectors of this space and operations over them are completely arithmetized in a unique manner (irrespective of the nature of objects constituting the given space). In other words all the n -dimensional linear spaces are isomorphic to one concrete n -dimensional linear space, the arithmetic space K_n .

§ 178. Suppose that linearly independent vectors $a_1, \dots, a_m$ are arbitrarily given in a linear space L . Consider the set L' of all linear combinations of vectors $a_1, a_2, \dots, a_m$, that is, the set of all vectors of the form

$$x = \lambda_1 a_1 + \lambda_2 a_2 + \dots + \lambda_m a_m$$

where $\lambda_1, \lambda_2, \dots, \lambda_m$ are any numbers. Obviously, if x and y are two vectors in L' , then $x + y$ is also in L' ; if λ is any number, then λx is in L' ; the null vector $\theta = 0 \cdot a_1 + 0 \cdot a_2 + \dots + 0 \cdot a_m$ and the vector $-x = (-\lambda_1) a_1 + \dots + (-\lambda_m) a_m$ belong to L' . Thus L' is a linear space in its own right. This space is isomorphic to the coordinate space K_m , and hence it is m -dimensional. The vectors $a_1, a_2, \dots, a_m$ constitute a basis for L' , and the numbers $\lambda_1, \lambda_2, \dots, \lambda_m$ are the coordinates of the vector x in L' with respect to the given basis.

§ 179. Let there be given a set $\mathfrak{A}$ whose elements will be henceforth called points and will be denoted by the capital latin letters A, B, C and an n -dimensional linear space L whose vectors will be denoted by the small latin letters $a, b, x, y, \dots$ (except for the null element, which is denoted by θ). Assume that with every ordered pair of points A, B in $\mathfrak{A}$ there is associated a vector x in L . If A is considered as the first point of the pair A, B and the vector x is associated with the pair A, B , then we shall write

$$AB = x$$

With any pair of the same points one vector of L is associated because this pair need not be considered as ordered. The association of vectors of L with the pairs of points of $\mathfrak{A}$ can be arbitrary but only the following two axioms are to be satisfied:

(1) *For any point A and any vector x there is a unique point B such that $AB = x$.*

(2) If $AB = x$, $BC = y$, then $AC = x + y$. The set of points which is associated with an n -dimensional linear space in the manner just described is called an n -dimensional affine space.

The following two theorems are simple consequences of Axioms 1, 2

(a) With every pair of coincident points the null vector is associated.

Indeed, let x be any vector and $AA = z$. By Axiom 1, there is a point B such that $AB = x$ and then Axiom 2 implies that $AB = z + x$. Therefore $z + x = x$, implying $z = \theta$.

(b) If $AB = x$, then $BA = -x$.

Indeed, if $BA = y$, Axiom 2 implies that $AA = x + y$, whence $x + y = \theta$ and $y = -x$.

In order to simplify the presentation, the ordered pairs of points of the affine space will henceforth be simply referred to as *vectors*. It is clear from what has been described that in the sense of linear operations (addition and multiplication by a number) the ordered pairs of points of the affine space play exactly the same role as the one played by ordinary geometrical vectors in the usual vector algebra. What is important, however, is that we are now discussing vectors in space of an arbitrary dimension.

§ 180. Affine coordinates may be introduced in the affine space. The affine coordinate system is defined by means of the origin O and the coordinate basis $e_1, e_2, \dots, e_n$. Taking any point M , we decompose its radius vector OM with respect to the coordinate basis:

$$OM = x_1 e_1 + x_2 e_2 + \dots + x_n e_n$$

The coefficients in this decomposition are called the *coordinates* of M in the given system.

If M^* is another point, then

$$OM^* = x_1^* e_1 + x_2^* e_2 + \dots + x_n^* e_n$$

whence

$$MM^* = OM^* - OM = (x_1^* - x_1) e_1 + \dots + (x_n^* - x_n) e_n$$

Thus, as in the usual vector algebra, the coordinates of the vector MM^* are obtained by subtracting the coordinates of M from those of M^* .

Let us derive the formulas for the transformation of affine coordinates when a new origin O' and a new basis $e'_1, e'_2, \dots, e'_n$ are taken. Suppose that the coordinates $(a_1, a_2, \dots, a_n)$ of O' in the old system and the coefficients in the decompositions of vectors of the new basis with

respect to the old basis are known:

$$\left. \begin{aligned} e'_1 &= p_{11}e_1 + p_{12}e_2 + \dots + p_{1n}e_n \\ e'_2 &= p_{21}e_1 + p_{22}e_2 + \dots + p_{2n}e_n \\ &\vdots \\ e'_n &= p_{n1}e_1 + p_{n2}e_2 + \dots + p_{nn}e_n \end{aligned} \right\} \quad (1)$$

Formulas (1) can be written in short form as

$$e'_i = \sum_{k=1}^n p_{ik} e_k \quad (i = 1, 2, \dots, n) \quad (2)$$

If M is any point, then

$$OM = x_1 e_1 + x_2 e_2 + \dots + x_n e_n = \sum_{k=1}^n x_k e_k$$

$$OM = O'M + OO' = \sum_{i=1}^n x_i e_i + \sum_{k=1}^n a_k e_k$$

(here primes denote the new coordinates of M). Comparing the right-hand sides in the last two relations and applying formula (2), we have

$$\begin{aligned} \sum_{k=1}^n x_k e_k &= \sum_{i=1}^n x'_i e'_i + \sum_{k=1}^n a_k e_k = \sum_{i=1}^n \left\{ x'_i \sum_{k=1}^n p_{ik} e_k \right\} + \sum_{k=1}^n a_k e_k \\ &= \sum_{k=1}^n \left\{ \sum_{i=1}^n p_{ik} x'_i + a_k \right\} e_k \end{aligned}$$

In this chain of identities the initial and final decompositions are both with respect to one and the same basis, and, because the decomposition relative to a basis is unique, we find that

$$x_k = \sum_{i=1}^n p_{ik} x'_i + a_k$$

or, in the expanded form,

[illegible]

These formulas express the old coordinates of any point in terms of its new coordinates. Let P be the matrix formed of the coefficients in the right side of formulas (4). Since the vectors $e'_1, \dots, e'_n$ are linearly independent, the rows of P must also be linearly independent. Accordingly, the determinant of P does not vanish. The determinant of matrix P^* , the transpose of P , does not vanish either, and there-

fore P^* has an inverse. Evidently, the coefficients of $x'_1, x'_2, \dots, x'_n$ in the right side of (3) are elements of P^* . If q_{ik} denote the elements of the matrix $Q = (P^*)^{-1}$, the inverse of P^* , then (3) yield

$$\left. \begin{aligned} x'_1 &= q_{11}x_1 + q_{12}x_2 + \dots + q_{1n}x_n + b_1 \\ x'_2 &= q_{21}x_1 + q_{22}x_2 + \dots + q_{2n}x_n + b_2 \\ &\vdots \\ x'_n &= q_{n1}x_1 + q_{n2}x_2 + \dots + q_{nn}x_n + b_n \end{aligned} \right\} \quad (4)$$

where $b_i = -\sum_{k=1}^n q_{ik}a_k$. Formulas (4) express the new coordinates of the point in terms of its old coordinates.

According to what has been said, if (4) represent a certain transformation of affine coordinates, the determinant of the matrix $Q = (q_{ik})$ does not vanish. Conversely, if formulas of form (4) are given beforehand with arbitrary numbers b_i and any matrix $Q = (q_{ik})$ with a nonvanishing determinant, they represent a transformation of affine coordinates. Indeed, if $\det Q \neq 0$, the a_h 's can be

found from the equations $\sum_{k=1}^n q_{ik} a_k = -b_i$, that is, the origin of the new system is found; furthermore, by inverting Q we find P^* , then P and finally the new basis from formulas (1).

§ 181. For the sake of definiteness, we shall henceforth consider the case $n = 4$. In the four-dimensional affine space, we can define the *lines*, *planes* and *hyperplanes* in a natural way.

Let A be a given point and a a given vector ($a \neq 0$). By a *line* through A in the direction of vector a we shall mean the set of points M satisfying the equation

$$AM = \lambda a \quad (1)$$

for all possible numerical values of the parameter λ ; the point A itself corresponds to the value $\lambda = 0$.

It is clear that all the points of the line enjoy equal privilege in the sense that any one of them can assume the role of A . Indeed, if B is any point of the line in question corresponding to the value $\lambda = \lambda_1$, then

$$BM = AM - AB = (\lambda - \lambda_1)a = \mu a \quad (2)$$

where $\mu = \lambda - \lambda_1$. Thus the set of points M determined by equation (1) with parameter λ can also be determined by equation (2) with parameter μ ; in view of (2), the point B corresponds to the value $\mu = 0$.

Suppose now that a point A and two linearly independent vectors a and b are given. By a *plane* passing through A in the direction of vectors a and b we shall mean the set of points M satisfying the

equation

$$AM = \lambda a + \mu b \quad (3)$$

for all possible numerical values of the parameters λ and μ .

Finally, if a point A and three linearly independent vectors a, b, c are given, the set of points M satisfying the equation

$$AM = \lambda a + \mu b + \nu c \quad (4)$$

for all possible numerical values of three parameters λ, μ, ν is called a *hyperplane* passing through A in the direction of the vectors a, b, c .

As in the case of a line, it may be easily demonstrated that all the points of a plane or a hyperplane enjoy equal importance in the sense that any one of them could assume the role of point A .

It is important to observe that a hyperplane can be treated as a three-dimensional affine space. Indeed, the set L' of all the linear combinations of the vectors a, b, c is a three-dimensional linear space (see § 178); at the same time if M_1 and M_2 are two points of the hyperplane determined from equation (4) with $\lambda = \lambda_1, \mu = \mu_1, \nu = \nu_1$ and $\lambda = \lambda_2, \mu = \mu_2, \nu = \nu_2$, then with the ordered pair of points M_1, M_2 there is associated a vector

$$M_1M_2 = (\lambda_2 - \lambda_1)a + (\mu_2 - \mu_1)b + (\nu_2 - \nu_1)c$$

of L' . This association satisfies the two axioms of § 179; accordingly, by the definition of § 179, the hyperplane is an affine space which is three-dimensional because so is the linear space L' . The parameters λ, μ, ν in equation (4) are nothing but the coordinates of M in the affine coordinate system which is defined on the hyperplane by taking A as the origin and the triple of vectors a, b, c as the basis. Of course, the same hyperplane may be represented by an equation of form (4) by taking any other point instead of A and any other three linearly independent vectors of L' instead of the vectors a, b, c ; such a change in equation (4) will amount to another affine coordinate system on the given hyperplane.

In the same manner one can show that every plane is a two-dimensional affine space and every line is a one-dimensional affine space.

§ 182. From the definition of lines, planes and hyperplanes, the following propositions can be easily proved.

(1) Given any two distinct points A and B , there is one and only one line passing through A and B (that is, containing these points); namely, this will be the line that passes through A in the direction of the vector $a = AB$.

(2) Given any three noncollinear points A, B, C , there is one and only one plane passing through the points A, B, C (namely, the plane passing through A in the directions of the vectors AB and AC).

(3) Given any four noncoplanar points A, B, C, D , there is one and only one hyperplane passing through the points A, B, C, D

(namely, the hyperplane passing through A in the directions of the vectors AB , AC , AD).

(4) If two distinct points A and B belong to one plane α , then all points of the line AB belong to α . To prove this, it is enough to represent α by the equation

$$AM = \lambda a + \mu b$$

by taking $a = AB$. Then all points of the line AB are determined by the same equation for different values of λ with $\mu = 0$.

(5) If two distinct planes α and β have two non-coincident points A , B , then all the common points of α and β lie on the line AB . In fact, if there were a point among the common points of α , β that does not lie on the line AB , the planes α , β must coincide, violating the assumption.

(6) If three noncollinear points A , B , C lie in a hyperplane α , then the entire plane ABC lies in α (the proof is similar to that of Proposition 4).

(7) If two distinct hyperplanes α , β have a common point, they intersect along a plane.

Proof. Let e_1, e_2, e_3 be linearly independent vectors in the hyperplane α . Since the hyperplanes α and β are distinct there is a vector e_4 in β such that e_1, e_2, e_3, e_4 are linearly independent. Moreover, there are two more vectors e_5, e_6 in β that constitute a linearly independent triple with e_4 . Because the space is four-dimensional, the vectors e_1, e_2, e_3, e_4, e_5 are connected by a linear equation

$$\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3 + \lambda_4 e_4 + \lambda_5 e_5 = 0$$

in which $\lambda_5 \neq 0$. Similarly,

$$\mu_1 e_1 + \mu_2 e_2 + \mu_3 e_3 + \mu_4 e_4 + \mu_6 e_6 = 0$$

where $\mu_6 \neq 0$. Set

$$a = \lambda_4 e_4 + \lambda_5 e_5 = -\lambda_1 e_1 - \lambda_2 e_2 - \lambda_3 e_3$$

$$b = \mu_4 e_4 + \mu_6 e_6 = -\mu_1 e_1 - \mu_2 e_2 - \mu_3 e_3$$

The vectors a and b belong to both α and β . On the other hand, these two vectors are linearly independent (because $\lambda_5 \neq 0$, $\mu_6 \neq 0$). Therefore, if A is a common point of α and β , then the equation

$$AM = \lambda a + \mu b$$

determines a plane belonging to α and β . This plane exhausts all the common points of the hyperplanes α and β , for otherwise α and β would coincide (in view of Proposition 3).

(8) If the plane α has a common point with the hyperplane β , then either α lies entirely in β or α and β intersect along a line (the proof is analogous to that of the preceding proposition).

§ 183. Let the equation $AM = \lambda a$ represent any line, and let M_1, M_2, M_3 be three distinct points of this line and $\lambda_1, \lambda_2, \lambda_3$ the corresponding values of the parameter λ . We shall say that M_2 lies between M_1 and M_3 if $\lambda_1 < \lambda_2 < \lambda_3$ or $\lambda_3 < \lambda_2 < \lambda_1$. If a vector $b = \sigma a$ ($\sigma \neq 0$) is taken instead of a , then the same line is represented by the equation $AM = \mu b$, where $\mu = \lambda/\sigma$. The points M_1, M_2, M_3 in the new equation correspond to the values $\mu_1 = \lambda_1/\sigma$, $\mu_2 = \lambda_2/\sigma$, $\mu_3 = \lambda_3/\sigma$ of the new parameter μ . It is clear that if λ_2 lies between λ_1 and λ_3 , then μ_2 lies between μ_1 and μ_3 . This shows that the stated definition does not depend on the choice of the direction vector of the line; it may be easily shown that this definition also does not depend on the choice of point A .

Having defined the "betweenness" concept, we can define a segment, a triangle, etc. For any triangle in a plane Pasch's axiom continues to hold; also true is the proposition that every line lying in a given plane divides the plane into two regions, etc.

§ 184. In the affine space, we can define in a natural way the concept of *parallelism* of two lines, of a line and a plane, etc. Two lines represented by the equations

$$A_1M = \lambda a_1, \quad A_2M = \lambda a_2$$

are called *parallel* if they do not coincide and if their direction vectors are proportional (that is, if a_2 is the product of a_1 and a number). The line

$$A_1M = \lambda a_1$$

is said to be parallel to the plane

$$A_2M = \lambda a_2 + \mu b_2$$

if the given line does not lie in the plane and if the vector a_1 can be expressed in terms of the vectors a_2, b_2 . The line

$$A_1M = \lambda a_1$$

is said to be parallel to the hyperplane

$$A_2M = \lambda a_2 + \mu b_2 + \nu c_2$$

if the given line does not lie in this hyperplane and if the vector a_1 may be expressed in terms of the vectors a_2, b_2, c_2 . Two planes

$$A_1M = \lambda a_1 + \mu b_1, \quad A_2M = \lambda a_2 + \mu b_2$$

are parallel if they do not coincide and if the vectors a_1, b_1 may be expressed in terms of the vectors a_2, b_2 . The plane

$$A_1M = \lambda a_1 + \mu b_1$$

is parallel to the hyperplane

$$A_2M = \lambda a_2 + \mu b_2 + \nu c_2$$

if the given plane does not lie in this hyperplane and if the vectors a_1, b_1 may be expressed in terms of the vectors a_2, b_2, c_2 . Finally two hyperplanes

$$A_1M = \lambda a_1 + \mu b_1 + \nu c_1, \quad A_2M = \lambda a_2 + \mu b_2 + \nu c_2$$

are parallel if they do not coincide and if the vectors a_1, b_1, c_1 may be expressed in terms of the vectors a_2, b_2, c_2 . The following propositions can be easily established.

(1) Two lines are parallel if and only if they lie in one plane and do not intersect; through every point not lying on a given line there passes one and only one line parallel to the given line.

(2) A line and a plane are parallel if and only if they lie in one hyperplane and do not intersect.

(3) A line is parallel to a hyperplane if and only if they do not intersect.

(4) A plane is parallel to a hyperplane if and only if they do not intersect.

(5) Two hyperplanes are parallel if and only if they do not intersect.

From the propositions just mentioned it is clear that the geometry of the three-dimensional affine space, at least that portion which is being developed in the present section, does not differ from the geometry presented in § 164 (see the Remark at the end of § 164).

§ 185. The assertions of the preceding article as well as those of § 183 may be easily proved algebraically (in exactly the same way as is done in the usual analytic geometry) if one makes use of the equations of geometric images in affine coordinates.

Let an affine coordinate system be given. Then every first degree equation

$$A_1x_1 + A_2x_2 + A_3x_3 + A_4x_4 + A_5 = 0$$

represents a hyperplane. Indeed, letting $(x_1^0, x_2^0, x_3^0, x_4^0)$ denote a solution of equation (1), this equation may be written in the form

$$A_1(x_1 - x_1^0) + A_2(x_2 - x_2^0) + A_3(x_3 - x_3^0) + A_4(x_4 - x_4^0) = 0$$

Setting $x_i - x_i^0 = u_i$, we have

$$A_1u_1 + A_2u_2 + A_3u_3 + A_4u_4 = 0$$

Equation (3) has three linearly independent solutions:

$$(a_1, a_2, a_3, a_4), \quad (b_1, b_2, b_3, b_4), \quad (c_1, c_2, c_3, c_4)$$

and every solution of equation (3) may be expressed as a linear combination of the following three equations:

$$x_i - x_i^0 = u_i = \lambda a_i + \mu b_i + \nu c_i \quad (i = 1, 2, 3, 4)$$

By assigning all possible numerical values to the parameters λ, μ, ν , we obtain all solutions u_i of equation (3), and hence all solutions x_i of equation (1). If M is the point with coordinates x_i and A the point with coordinates x_i^0 and a, b, c are the vectors with coordinates a_i, b_i, c_i , then the scalar equations (4) are equivalent to one vector equation

$$AM = \lambda a + \mu b + \nu c \quad (5)$$

Hence we have shown that the set of points M the coordinates of which satisfy equation (1) coincides with the hyperplane represented by equation (5).

Conversely, every hyperplane is represented by a first-degree equation of form (1). Indeed, if the hyperplane is given by an equation of form (5), then taking the equivalent equations (4) and eliminating the parameters λ, μ, ν , we obtain an equation of form (2), which is easily reduced to form (1).

From the assertion just established and Propositions 7 and 8 of § 182 it follows that (a) two consistent and independent first-degree equations determine a plane, and (b) three consistent and independent first-degree equations determine a line.

§ 186. In the affine space we can examine hypersurfaces of the second order, that is, hypersurfaces which are represented by a second degree equation in affine coordinates. We shall not deal with the classification of second-order hypersurfaces; in general appearance it is analogous to the well known affine classification of second-order surfaces in the three-dimensional space. We shall confine ourselves to a particular case, important for later.

Let (x_1, x_2, x_3, x_4) be the coordinates of a variable point M and $(x_1^0, x_2^0, x_3^0, x_4^0)$ the coordinates of a fixed point A . Consider the equation

$$\sum_{i, k=1}^4 g_{ik} (x_i - x_i^0) (x_k - x_k^0) = 0 \quad (1)$$

the left side of which is a quadratic form in the arguments $x_1 - x_1^0, \dots, x_4 - x_4^0$ with coefficients g_{ik} ; this form will be denoted Φ . If we set $x_i = x_i^0$, equation (1) is satisfied, which means that A lies on the hypersurface represented by equation (1). Let M be another point whose coordinates satisfy equation (1). We shall displace M along the line originating from A . Then the difference $x_i - x_i^0$ will undergo a proportional change and the left side of equation (1) will continue to vanish. Consequently, if a point M lies on the hypersurface (1), all the points of the line AM lie on this hypersurface. The hypersurface (1) is thus composed of lines passing through A , and is therefore called a quadric cone with vertex A . Of course, it might happen that there is no point, except A , which satisfies equation (1); this case occurs if Φ is a form having a fixed sign. In this case, the

cone is called *imaginary*. If, however, Φ changes its sign and is a nondegenerate form (that is, a form with variable sign whose determinant $\det g_{ik} \neq 0$), then the cone (a) possesses infinitely many lines as its generators, (b) is four-dimensional, that is, does not lie entirely in any one hyperplane, (c) divides the space into two regions for one of which $\Phi > 0$ and for the other $\Phi < 0$. Such a cone is called a *real nondegenerate quadric cone*. We refrain from proving that a real nondegenerate cone has the properties just mentioned, and explain the situation by an example. Consider the equation

$$(x_1 - x_1^0)^2 + (x_2 - x_2^0)^2 + (x_3 - x_3^0)^2 - (x_4 - x_4^0)^2 = 0 \quad (2)$$

the left side of which is a quadratic form in $x_i - x_i^0$ with coefficients $g_{11} = g_{22} = g_{33} = 1$, $g_{44} = -1$, $g_{ik} = 0$, ($i \neq k$). This form is nondegenerate because $\det g_{ik} = -1 \neq 0$ and is of variable sign (the form is positive if $x_1 \neq x_1^0$, $x_2 \neq x_2^0$, $x_3 = x_3^0$, $x_4 = x_4^0$ and negative if $x_1 = x_1^0$, $x_2 = x_2^0$, $x_3 = x_3^0$, $x_4 \neq x_4^0$). Consequently, equation (2) represents a real nondegenerate quadric cone. In order to clearly visualize the properties of the cone represented by equation (2), it is beneficial to observe that any hypersurface $x_4 - x_4^0 = C$ cuts this cone along the three-dimensional sphere

$$(x_1 - x_1^0)^2 + (x_2 - x_2^0)^2 + (x_3 - x_3^0)^2 = C^2$$

in the same way as a plane perpendicular to the axis of the usual circular cone cuts the cone in a circle. The set of points for which the left side of equation (2) is negative is called the *interior region of the cone* (2). The interior region is divided into two sheets for one of which $x_4 > x_4^0$ and for the other $x_4 < x_4^0$.

§ 187. Now we shall prove a proposition which will play an especially important role below. Let an affine coordinate system be given, and let the four functions

$$\left. \begin{aligned} x'_1 &= f_1(x_1, x_2, x_3, x_4) \\ x'_2 &= f_2(x_1, x_2, x_3, x_4) \\ x'_3 &= f_3(x_1, x_2, x_3, x_4) \\ x'_4 &= f_4(x_1, x_2, x_3, x_4) \end{aligned} \right\} \quad (1)$$

be given in which each of the functions is defined in the entire space; this amounts to defining a mapping of the space into itself because to every point $M(x_1, x_2, x_3, x_4)$ there is assigned a point $M'(x'_1, x'_2, x'_3, x'_4)$. Assume that the mapping (1) is a one-to-one mapping of the space into itself (to every image point M' there corresponds just one preimage M); assume further that the mapping is a collineation, that is, for any three collinear points M_1, M_2, M_3 the corresponding images M'_1, M'_2, M'_3 are also collinear. Under these assumptions the functions f_1, f_2, f_3, f_4 are linear, that is, they are of the

form

$$\left. \begin{aligned} x'_1 &= q_{11}x_1 + q_{12}x_2 + q_{13}x_3 + q_{14}x_4 + b_1 \\ x'_2 &= q_{21}x_1 + q_{22}x_2 + q_{23}x_3 + q_{24}x_4 + b_2 \\ x'_3 &= q_{31}x_1 + q_{32}x_2 + q_{33}x_3 + q_{34}x_4 + b_3 \\ x'_4 &= q_{41}x_1 + q_{42}x_2 + q_{43}x_3 + q_{44}x_4 + b_4 \end{aligned} \right\} \quad (2)$$

and the determinant of the matrix $Q = (q_{ik})$ does not vanish. In short, if a mapping is a collineation it is linear.

We shall present the main steps of the proof omitting some minor details in the process.

(1) Let us examine a one-to-one collineation of the affine space into itself. Let A, B, C be three noncollinear points in space and α the plane determined by A, B, C . Assume that the images A', B', C' of these points are also noncollinear and let α' be the plane through A', B', C' . Then the images of all points in α lie in α' . Indeed, taking any point M in α , we draw through M two distinct lines a and b in α such that a meets the lines AC and BC in two distinct points P, Q and b meets the lines AC and BC in two distinct points R, S . The mapping in question (and Proposition 4, § 182) imply that the points P', Q', R', S' , which are images of P, Q, R, S under the given mapping, lie in α' . But the image M' of M is determined by the intersection of the lines $P'Q'$ and $R'S'$; accordingly, M' also lies in α' .

(2) According to the definition of a collineation, the images of points of a line a lie on a definite line a' ; we shall say that a' corresponds to a under the given mapping. If any two lines a, b lying in the plane α , which figured in the preceding item, are parallel, then their corresponding lines a', b' are parallel in α' (because the mapping is one-to-one). Therefore we may adjoin points at infinity to the planes α and α' (in the same way as we did in §§ 80 and 81) and extend the given mapping to the augmented planes by assuming that the image of the point at infinity on a in α is the point at infinity on a' in α' . Thus to every point of α there corresponds a point of α' ; to the collinear points in α there correspond points in α' that are also collinear; to the line at infinity in α there corresponds the line at infinity in α' ; finally, there are three noncollinear points A, B, C in α the images A', B', C' of which are also noncollinear in α' . By § 106, such a mapping is a projectivity of the augmented plane α into the augmented plane α' .

(3) Assume now that under the given mapping the points A, B, C retain their positions, that is, they coincide with their images. In that case the plane α retains its position and is mapped projectively into itself. Notice that the points at infinity of the lines CA and CB also remain fixed, together with the points A, B, C . Denote by a the line passing through A and the point at infinity of the line CB , and by b the line through B and the point at infinity of the line CA . The

lines a and b remain fixed; accordingly, the point P , where these lines meet, also remains fixed. Thus four points A, B, C, P in α remain fixed and no three of these points are collinear. This in conjunction with Theorem 26, § 106, implies that all the points in α remain fixed.

(4) Assume that A, B, C, D are four noncoplanar points in space which remain fixed under the given mapping of the space. Let α denote the hyperplane determined by A, B, C, D ; we shall show that all the points of α remain fixed. Consider any point M in α . Let K be the point of intersection of the line DM and the plane ABC ; the results of the preceding item imply that K remains fixed. Similarly, the point of intersection of the line CM and the plane ABD also remains fixed, which implies that M itself remains fixed.

(5) Let A, B, C, D, E be five points in space that do not lie in one hyperplane. If these points remain fixed, then all points in space remain fixed. This assertion can be derived from that of item (4) in the same way as the result of item (4) was derived from that of item (3).

(6) Consider now a collineation as given in the theorem; we shall represent this mapping by $M' = f(M)$. Let O, e_1, e_2, e_3, e_4 be the origin and basis, respectively, of the given affine coordinate system and let A_1, A_2, A_3, A_4 be the endpoints of the basis vectors which have been affixed at O . The definition of a basis implies that the five points O, A_1, A_2, A_3, A_4 do not lie in one hyperplane; in that case their preimages $O^*, A_1^*, A_2^*, A_3^*, A_4^*$ also do not lie in one hyperplane. Therefore there exists an affine coordinate system with origin O^* and the basis consisting of the vectors $a_i^* = O^*A_i^*$ ($i = 1, 2, 3, 4$). We shall call this system the new system and the system given initially the old system. Let M be any point in the space, x_i^* its coordinates in the new system, M' the image of M under the given mapping, and x_i the coordinates of M' in the old system. We define another mapping $M'' = \varphi(M)$ as follows: the point M'' has the same coordinates x_i^* in the old system as M has in the new system. The mapping $M'' = \varphi(M)$ is one-to-one, and, moreover, both this mapping and its inverse $M = \psi(M'')$ are collineations (indeed, if for example M moves along a line which is given by three first-degree equations in the new system, then the trajectory of M'' is defined by exactly the same equations in the old system and is, accordingly, also a line). Notice that $M = \psi(M'')$ maps the points O, A_i into the points O^*, A_i^* . We now construct the mapping $M' = f(\psi(M''))$; in other words, first the point M'' is mapped into the point $M = \psi(M'')$ and then M into $M' = f(M)$. The mapping $M' = f(\psi(M''))$ is one-to-one and a collineation (because the component mapping are too); furthermore, $M' = f(\psi(M''))$ leaves O, A_i fixed (because first these points go to O^*, A_i^* and then return to their positions). This and the results of item (5) imply that all the points remain fixed

under the mapping $M' = f(\psi(M''))$, that is, every point M' coincides with its preimage M'' . Hence the point $M' = f(M)$ has exactly the same coordinates in the old system as M has in the new system: $x'_i = x_i^*$. But according to § 180, the new coordinates x_i^* of any point are linear combinations of its old coordinates x_i . Hence x'_i are linear functions of x_i , that is, are of form (2). That the determinant of matrix Q does not vanish follows from the fact that formula (2) is uniquely invertible in view of the hypothesis of the theorem. ■

§ 188. The one-to-one collineation of the affine space into itself is called the *affine transformation* in this space. The theorem just proved states that every affine transformation can be expressed in affine coordinates by linear formulas of form (2) with a non-vanishing determinant of the matrix Q . Every affine transformation has an inverse which is also an affine transformation; this follows from the theorem proved (because a collineation is represented by linear formulas, its inverse is also represented by linear formulas and is hence a collineation). It is also evident, furthermore, that the product of two affine transformations is also an affine transformation. Therefore all the affine transformations of the affine space constitute a group; this group is referred to as the *affine group* of this space. The theory of invariants for the affine group of the n -dimensional affine space is called the *n -dimensional affine geometry*. The concepts of a line, plane, hyperplane, parallelism, etc. are invariants with respect to the affine group, and, accordingly, they are objects of affine geometry.

7.2. Euclidean spaces and Minkowski space

§ 189. Consider an n -dimensional (real) affine space $\mathfrak{A}$. Assume that with every pair of vectors x, y of this space there is associated a real number, denoted by xy , satisfying the following three axioms:

- (1) $xy = yx$.
- (2) $x(\lambda y + \mu z) = \lambda(xy) + \mu(xz)$, where λ, μ are any real numbers.

These two axioms imply, in particular, that $\theta x = 0$ for the null vector θ and any vector x (since $\theta = 0 \cdot x$, it follows that $\theta x = x\theta = = x(0 \cdot x) = 0(x) = 0$).

- (3) If $xy = 0$ for a certain x and any y , then $x = \theta$.

The quantity xy is called the *scalar product* of the vector x and y . The n -dimensional affine space endowed with a scalar product of its vectors is called an *n -dimensional Euclidean space* (we have defined a real Euclidean space, since $\mathfrak{A}$ was assumed to be a real affine space and xy a real number).

§ 190. Taking some n -dimensional Euclidean space, we choose in this space a system of affine coordinates with O as the origin and

$e_1, e_2, \dots, e_n$ the basis of this system. Denoting the scalar product of any pair of basis vectors e_i, e_k by g_{ik} :

$$e_i e_k = g_{ik} \quad (1)$$

we find that by virtue of Axiom 1 we must have $g_{ik} = g_{ki}$. Let now x, y be any vectors and

$$x = \sum_{i=1}^n x_i e_i, \quad y = \sum_{k=1}^n y_k e_k \quad (2)$$

be their decompositions with respect to the given basis. Take the scalar multiplication of the left and right sides of (2) and multiply the right sides termwise (this is permissible by Axiom 2). Noting the multiplication table (1) for the basis vectors, we have

$$xy = \sum_{i,k=1}^n g_{ik} x_i y_k \quad (3)$$

the right side of which is a bilinear form in coordinates of the vectors x, y with coefficients g_{ik} .

If g denotes the determinant of the matrix (g_{ik}) , Axiom 3 implies that $g \neq 0$ (that is, (g_{ik}) is a nondegenerate matrix).

Indeed, if $g = 0$, then a vector $x \neq \theta$ can be chosen such that for its coordinates x_i all the sums $\sum_{i=1}^n g_{ik} x_i$ ($k = 1, 2, \dots, n$) will be zero, but then $xy = 0$ for any y , which is ruled out by Axiom 3.

Thus, in an n -dimensional Euclidean space the scalar product xy is expressed by a bilinear form in the coordinates of vectors x, y and the coefficients of this form constitute a nondegenerate symmetric matrix.

Suppose now that an n -dimensional affine space is given and we wish to introduce a scalar product in this space, that is, we wish to turn this space into a Euclidean space. To this end we choose an affine coordinate system in the given space, choose the numbers g_{ik} satisfying the condition $g_{ik} = g_{ki}$ and with any pair of vectors x, y associate the number xy according to formula (3). Axioms 1 and 2 will then be fulfilled because the matrix (g_{ik}) chosen is symmetric and the right side of (3) is linear in x_i and y_i . In order to satisfy Axiom 3, it is necessary to choose the quantities g_{ik} such that the determinant g of the matrix (g_{ik}) does not vanish. It is easy to see that this condition is sufficient as well. In fact, assume that $g \neq 0$. If

$xy = 0$ for any y , then (3) implies $\sum_{i=1}^n g_{ik} x_i = 0$ ($k = 1, 2, \dots, n$) and because $g \neq 0$, these equations imply that $x_i = 0$ or that $x = \theta$.

Thus, if in an n -dimensional affine space the quantity xy is defined in accordance with (3) by taking in the right side any bilinear form

with a nondegenerate symmetric matrix, xy will satisfy all the axioms for a scalar product.

Remark. As shown just now, Axiom 3 is equivalent to nondegeneracy of the matrix (g_{ik}) . On this account Axiom 3 is called the *nondegeneracy condition*.

§ 191. The following important concepts are introduced in the Euclidean space.

(1) *The orthogonality of vectors, lines, etc.* Two vectors x and y are called *orthogonal*, or *perpendicular*, if $xy = 0$. Two lines are orthogonal if their direction vectors are orthogonal; a line and a plane are orthogonal if the direction vector of the line is orthogonal to every direction vector of the plane; the orthogonality of a line and a hyperplane is similarly defined.

(2) *The norm of a vector.* The norm of a vector x , denoted by $\|x\|$, is defined by

$$\|x\| = \sqrt{x^2} \quad (4)$$

where $x^2 = xx$. For the sake of definiteness we shall take the positive square root. However, it should be noted that the definition of a scalar product given above does not rule out the case $x^2 < 0$; in this case the vector has an imaginary norm. The case $\|x\| = 0$ for $x \neq 0$ is not ruled out either.

The vector x is called a *unit vector* if $x^2 = 1$, *imaginary-unit* if $x^2 = -1$, and *isotropic* if $x^2 = 0$ at $x \neq 0$.

Formulas (3) of § 190 and (1) of § 191 imply

$$\|x\|^2 = \sum_{i,k=1}^n g_{ik} x_i x_k \quad (2)$$

expressing the norm of a vector in terms of the coordinates. On the right side there appears a quadratic form in the coordinates of the vector x ; this form is called the *metric form of the Euclidean space*. Because $\det g_{ik} \neq 0$, the metric form is nondegenerate.

(3) *The distance between two points.* The distance between two points A and B is taken to be the norm of the vector AB :

$$\rho(A, B) = \|AB\|$$

We shall denote the coordinates of points by capital letters (in order not to confuse them with the coordinates of vectors). Suppose that the coordinates of A and B are $(X_1, X_2, \dots, X_n)$ and $(X_1^*, X_2^*, \dots, X_n^*)$. Then the coordinates of the vector AB are $x_1 = X_1^* - X_1$, $x_2 = X_2^* - X_2$, etc.; therefore by formula (2) we have

$$\rho^2(A, B) = \sum_{i,k=1}^n g_{ik} (X_i^* - X_i) (X_k^* - X_k) \quad (3)$$

The general definition of the Euclidean space does not rule out the possibilities that the distance between some points may be imaginary or zero. Take a fixed point A with coordinates X_i^0 and a variable point M with coordinates X_i . We shall find all the points M whose distance from A is zero; for the coordinates of these points we have the equation

$$\sum_{i,k=1}^n g_{ik} (X_i - X_i^0) (X_k - X_k^0) = 0 \quad (4)$$

If the metric form has a definite sign, then equation (4) is satisfied only for $X_i = X_i^0$, and in this case $\rho(A, M) = 0$ only if M coincides with A . If, however, the metric form has a variable sign, then equation (4) determines a real nondegenerate quadric cone with vertex A ; this cone is called the *isotropic cone at A* (the isotropic cone is nondegenerate because the metric form is nondegenerate; see § 186). The lines generating the isotropic cone are known as the *isotropic lines*. Every isotropic line is characterized by the fact that the distance between any pair of its points is zero.

§ 192. In the affine space every line, plane or hyperplane is an affine space of corresponding dimension in its own right (see § 181). If the affine space has been turned into a Euclidean space, that is, for every pair of its vectors the scalar product is defined, then by the same token the scalar product for every pair of vectors of the given line, plane or hyperplane is defined. Accordingly, the given line, plane or hyperplane turns into a Euclidean space of corresponding dimension provided on this line, plane or hyperplane the nondegeneracy condition is satisfied, which may not always be the case. Namely, if $xy = 0$ for a certain x and any y , then the nondegeneracy condition implies that $x = 0$, but it might happen that on some line, plane or hyperplane there is a vector $x \neq 0$ such that $xy = 0$ for any y lying on this line, plane or hyperplane. For example, the scalar product of any two vectors lying on the isotropic line is zero. Just like the lines, the planes and hyperplanes of the Euclidean space on which the nondegeneracy condition fails to be satisfied are called *isotropic*.

§ 193. Suppose that $e_1, \dots, e_n$ is the basis for an affine coordinate system in which the metric form in space is of the form (2), § 191. We take a new basis $e'_1, \dots, e'_n$ according to the formula

$$e'_i = \sum_{k=1}^n p_{ik} e_k, \quad i = 1, 2, \dots, n, \quad \det p_{ik} \neq 0 \quad (1)$$

Then the old coordinates x_i of any vector x are expressed in terms of the new coordinates x'_i thus

$$x_k = \sum_{i=1}^n p_{ik} x'_i \quad k = 1, 2, \dots, n \quad (2)$$

(See § 180; the location of the new and old origins is immaterial because we are now considering coordinates of vectors and not of points; the formulas for transformation of coordinates of vectors are homogeneous, that is, the free terms in the right sides of these formulas are zeroes.) Replacement in metric form (2), § 191, of $x_1, \dots, x_n$ with their expressions in terms of $x'_1, \dots, x'_n$ gives the metric form in the new coordinates. According to the theory of quadratic forms the coefficients p_{ih} in the linear transformation (2) can be chosen (with $\det p_{ih} \neq 0$) such that in the new coordinates the metric form is in the canonical form, that is, this contains only the squares of the coordinates, the number of terms is exactly n (because the form is nondegenerate) and the coefficients of the terms are either $+1$ or -1 . In other words, if σ_{ih} denote the coefficients of the transformed form, then

$$\begin{aligned}\sigma_{ii} &= \pm 1, & i &= 1, 2, \dots, n, \\ \sigma_{ih} &= 0, & i &\neq h, \quad i, h = 1, 2, \dots, n\end{aligned}$$

In these special coordinates,

$$\|x\|^2 = x_1'^2 + \dots + x_m'^2 - x_{m+1}'^2 - \dots - x_n'^2$$

Here the first m coefficients σ_{ii} are positive, which can always be achieved by redesignating the coordinates. The cases when all the coefficients are positive ($m = n$) and when all are negative ($m = 0$) are not ruled out. Taking note of the fact that $e'_i e'_k = \sigma_{ih}$, we have

$$\begin{aligned}e_1'^2 = \dots = e_m'^2 &= +1, & e_{m+1}'^2 = \dots = e_n'^2 &= -1 \\ e'_i e'_k &= 0, & i &\neq k\end{aligned}$$

that is, the basis vectors are unit or imaginary-unit and are pairwise orthogonal. Such a basis is called *orthonormalized*. We have proved that an orthonormal basis exists in every Euclidean space.

The reduction of the quadratic form to a canonical form can be effected in an infinite number of ways; this means that there are infinitely many distinct orthonormal bases exist in a Euclidean space. According to the law of inertia of the theory of quadratic forms the number of negative terms in the canonical representation of a metric form is independent of the manner of reduction to canonical form. This number expresses the geometrical properties of the given Euclidean space and is known as the *index* of the space. At the same time, the index is the number of imaginary-unit vectors in any orthonormal basis.

If the index is zero, then the norm of a vector, the scalar product of two vectors, etc. are represented by formulas that are totally analogous to the well known formulas of the usual analytic geometry. In this case the geometrical properties of the space are essentially the same as those of the usual three-dimensional Euclidean space;

speaking precisely, they may differ only in dimension. On this account, the Euclidean space with zero index is called *proper Euclidean* and the remaining Euclidean spaces are called *pseudo-Euclidean*. The Euclidean space with index 1 is known as the *Minkowski space*, and this will be the subject of our further discussion.

§ 194. Assume that a coordinate system with an origin O and orthonormal basis $e_1, \dots, e_n$ has been introduced in the Minkowski space. The basis vectors are assumed to have been arranged in such a way that $e_1^2 = \dots = e_{n-1}^2 = +1$, $e_n^2 = -1$. Then the norm of a vector x with the coordinates x_i is given by

$$\|x\|^2 = x_1^2 + \dots + x_{n-1}^2 - x_n^2 \quad (1)$$

the scalar product of two vectors x, y with the coordinates x_i, y_i by

$$xy = x_1y_1 + \dots + x_{n-1}y_{n-1} - x_ny_n \quad (2)$$

and the square of distance between two points $A(X_i), B(X_i^*)$ by

$$\rho^2(A, B) = (X_1^* - X_1)^2 + \dots + (X_{n-1}^* - X_{n-1})^2 - (X_n^* - X_n)^2 \quad (3)$$

An isotropic cone with vertex $A(X_1^0, \dots, X_n^0)$ is represented in the given coordinates by the equation

$$(X_1 - X_1^0)^2 + \dots + (X_{n-1} - X_{n-1}^0)^2 - (X_n - X_n^0)^2 = 0 \quad (4)$$

and is real and nondegenerate. The points where the left side of (4) is negative constitute the interior of the isotropic cone; the interior of the cone is divided into two sheets for one of which $X_n > X_n^0$ and for the other $X_n < X_n^0$.

§ 195. For visualization, let us consider a few objects of the Minkowski space when $n = 2$ and $n = 3$.

(1) Let us construct a model for the two-dimensional Minkowskian geometry on the Euclidean plane. First of all, the points, vectors and linear operations over these vectors are assumed to have their usual significance. Choose an affine coordinate system with origin O and basis e_1, e_2 ; the coordinates X_1, X_2 of any point M will also have their usual meaning (for example, X_1 is represented by a segment which intersects the line passing through M and parallel to the second axis; of course, this segment must be measured by the scale e_1). What is more, we are perfectly justified in taking the vectors e_1, e_2 of the same length and perpendicular in the Euclidean sense. Then the coordinate system will be simply the rectangular Cartesian system. However, the scalar product of two vectors x, y with the coordinates x_i, y_i ($i = 1, 2$) is defined in the sense of Minkowskian geometry by

$$xy = x_1y_1 - x_2y_2$$

and the norm of vector x by

$$\|x\|^2 = x_1^2 - x_2^2$$

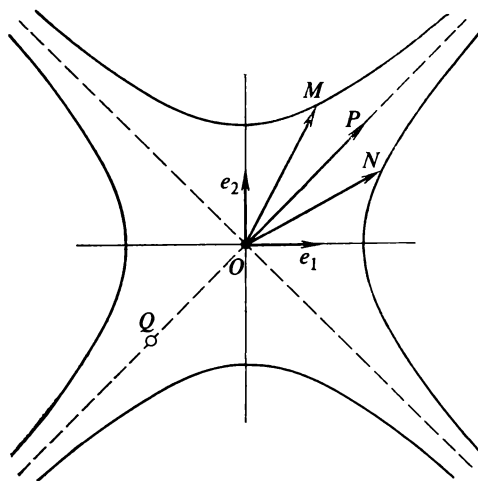


Figure 158

The isotropic cone, whose vertex is taken at the origin for simplicity, is given by the equation

$$X_1^2 - X_2^2 = 0$$

This cone consists of two Euclidean coordinate bisectors (Fig. 158). The norm of vector OP on any one of these bisectors is zero; the distance between any two points P, Q of the coordinate bisector is zero. The interior of the isotropic cone is defined by the inequality $X_1^2 - X_2^2 < 0$; this is composed of points lying in the interior of two vertical angles one of which is bounded by the upper rays of the bisector and the other by the lower rays. The distance between any point M in the interior of the isotropic cone and the origin is imaginary. Let $\rho(O, M) = ai$; then all the points which are at the same distance from O satisfy the equation

$$X_1^2 - X_2^2 = -a^2$$

In the sense of Minkowskian geometry, these points lie on a circle of imaginary radius ai , while in the Euclidean sense they lie on the usual hyperbola (because the last equation represents a hyperbola with vertices on the second coordinate axis). The distance between any point N lying in the exterior of the isotropic cone and the origin O is a real quantity. Let $\rho(O, N) = a$; then all the points lying at the same distance from O satisfy the equation

$$X_1^2 - X_2^2 = a^2$$

In the Euclidean sense this equation represents a hyperbola with vertices on the first coordinate axis, while in the sense of Minkowskian geometry the same hyperbola represents a circle with real radius a .

The vectors OM and ON with coordinates (x_1, x_2) , (y_1, y_2) are perpendicular in the sense of Minkowski if $x_1y_1 - x_2y_2 = 0$; in the Euclidean sense this relation expresses the symmetry of the directions of OM and ON with respect to the coordinate bisectors. In particular, two vectors lying on one coordinate bisector are perpendicular in the sense of Minkowskian geometry.

(2) A model for the three-dimensional Minkowskian geometry can be constructed along the same lines as in the preceding case in the three-dimensional Euclidean space. Take the usual rectangular coordinate system with the basis e_1, e_2, e_3 . For any two vectors x, y with coordinates x_i, y_i ($i = 1, 2, 3$) we define the scalar product in the sense of Minkowski by putting

$$xy = x_1y_1 + x_2y_2 - x_3y_3$$

Then the norm of vector x is given by

$$\|x\|^2 = x_1^2 + x_2^2 - x_3^2$$

the isotropic cone by

$$X_1^2 + X_2^2 - X_3^2 = 0$$

and its interior by

$$X_1^2 + X_2^2 - X_3^2 < 0$$

From the Euclidean point of view, the isotropic cone is the usual cone of rotation about the third coordinate axis; its interior consists of two sheets of this cone. Any point lying in the interior of the isotropic cone lies at an imaginary distance from the origin. If this distance is called ai , then all the points lying at the same distance from O constitute a sphere of imaginary radius ai , and in the Euclidean sense this is a hyperboloid of two sheets whose equation is

$$X_1^2 + X_2^2 - X_3^2 = -a^2$$

The Euclidean hyperboloid of one sheet having the equation

$$X_1^2 + X_2^2 - X_3^2 = a^2$$

represents a sphere of real radius a in Minkowskian geometry.

If in the formulas expressing xy and $\|x\|^2$ the third coordinates are set equal to zero, then one obtains the two-dimensional formulas of the usual vector algebra. This means that on the coordinate plane passing through the vectors e_1, e_2 proper Euclidean geometry holds. In general, any plane that passes through the origin and does not contain any generator of the isotropic cone is a two-dimensional proper Euclidean space (because it does not contain isotropic lines).

Any plane passing through the origin and cutting the isotropic cone along two generators is a two-dimensional Minkowski space. The isotropic cone of Minkowski metric on this plane and the interior of this cone are determined by the intersection of the plane and the spatial isotropic cone. If the plane cutting the spatial isotropic cone turns into a tangent plane, then the lines generating the isotropic cone of the plane coincide and the interior of the cone vanishes. The isotropic cone of such a plane is degenerate. Hence, every plane touching the isotropic cone of the three-dimensional Minkowski space is an isotropic plane of this space.

The properties of a four-dimensional Minkowski space can be developed in a natural analogy with the three-dimensional model just examined.

§ 196. An affine transformation of the Minkowski space under which the distance between any two points is the same as that between their images will be called a *motion* in this space. Suppose that in a four-dimensional Minkowski space an affine coordinate system is introduced with origin O and an orthonormal basis e_1, e_2, e_3, e_4 ($e_i^2 = -1$). Then every affine transformation mapping a point $M(X_i)$ into a point $M'(X'_i)$ can be represented by formulas of form (2). § 187, this can be written in short as

$$X'_i = \sum_{k=1}^4 q_{ik} X_k + b_i, \quad i = 1, 2, 3, 4 \quad (1)$$

It is easily understood that transformation (1) in general will not preserve the distance between points; to do so the coefficients must satisfy some conditions which will be obtained now.

Let us discuss first a particular case of (1):

$$X'_i = X_i + b_i \quad (2)$$

Suppose that the points $M(X_i)$ and $N(X_i^*)$ are mapped by (2) into the points $M'(X'_i)$ and $N(X_i^{*'})$. If $x_i = X_i^* - X_i$ are the coordinates of vector MN and $x'_i = X_i^{*'} - X'_i$ those of vector $M'N'$, then (2) implies $x'_i = x_i$ and hence the norms of the vectors MN and $M'N'$ are equal. Thus for any b_i the transformation (2) preserves the distance between points; this particular case of a motion is known as *parallel translation*. Evidently, the parallel translation can be so chosen that any preassigned point becomes the origin.

In another particular case of (1), where $b_i = 0$, the origin remains fixed.

Any transformation of form (1) can be obtained by a chain of the two transformations just examined: first the origin and the basis are moved into a new position by the parallel translation and then another transformation is applied which is expressed by formulas (1) in new coordinates with the same matrix (q_{ik}) but in which the free

terms are zeroes. Because the parallel translation is clearly of no interest, formulas (1) will henceforth be assumed to be homogeneous:

$$X'_i = \sum_{k=1}^4 q_{ik} X_k \quad (3)$$

Let $M(X_i)$ and $N(X_i^*)$ be any two points, $M'(X'_i)$ and $N'(X'^*_i)$ their images, and let $x_i = X_i^* - X_i$ and $x'_i = X'^*_i - X'_i$ be the coordinates of the vectors MN and $M'N'$. Formulas (3) yield

$$x'_i = \sum_{k=1}^4 q_{ik} x_k \quad (4)$$

The equality of distances $\rho(M', N')$ and $\rho(M, N)$ is equivalent to that of the norms of vectors $M'N'$ and MN . Consequently, formulas (3) define a motion if the coefficients q_{ik} are chosen such that

$$x'^2_1 + x'^2_2 + x'^2_3 - x'^2_4 = x^2_1 + x^2_2 + x^2_3 - x^2_4 \quad (5)$$

The relation (5) must be fulfilled as a consequence of (4) for any x_1, x_2, x_3, x_4 . Denoting by σ_{ik} the coefficients of the metric form in orthonormalized coordinates ($\sigma_{11} = \sigma_{22} = \sigma_{33} = +1$, $\sigma_{44} = -1$, $\sigma_{ik} = 0$ if $i \neq k$), we write the left hand side of (5) as a double sum with coefficients σ_{ik} , making use of formula (4),

$$\begin{aligned} x'^2_1 + x'^2_2 + x'^2_3 - x'^2_4 &= \sum_{i,k=1}^4 \sigma_{ik} x'_i x'_k = \sum_{i,k=1}^4 \sigma_{ik} \left(\sum_{\alpha=1}^4 q_{i\alpha} x_\alpha \right) \left(\sum_{\beta=1}^4 q_{k\beta} x_\beta \right) \\ &= \sum_{\alpha,\beta=1}^4 \left(\sum_{i,k=1}^4 \sigma_{ik} q_{i\alpha} q_{k\beta} \right) x_\alpha x_\beta \end{aligned}$$

The right hand of (5) can also be written as a double sum:

$$x^2_1 + x^2_2 + x^2_3 - x^2_4 = \sum_{\alpha,\beta=1}^4 \sigma_{\alpha\beta} x_\alpha x_\beta$$

By virtue of (5), the expressions obtained must be the same, whence

$$\sum_{i,k=1}^4 \sigma_{ik} q_{i\alpha} q_{k\beta} = \sigma_{\alpha\beta} \quad (\alpha, \beta = 1, 2, 3, 4) \quad (6)$$

which is the desired condition on the coefficients q_{ik} for the transformation (3) or (4) to be a motion. The conditions (6) can be written in the matrix form. As before, let Q be the matrix with elements q_{ik} , the matrix Q^* with elements $q^*_{\alpha i} = q_{i\alpha}$ (Q^* is the transpose of Q), and I the matrix with elements σ_{ik} . Then relations (6) can be written as

$$\sum_{i,k=1}^4 q^*_{\alpha i} \sigma_{ik} q_{k\beta} = \sigma_{\alpha\beta} \quad (\alpha, \beta = 1, 2, 3, 4) \quad (7)$$

In this form they are evidently equivalent to the single matrix equation

$$Q^*IQ = I \quad (8)$$

or, in the expanded form,

$$\begin{pmatrix} q_{11} & q_{21} & q_{31} & q_{41} \\ q_{12} & q_{22} & q_{32} & q_{42} \\ q_{13} & q_{23} & q_{33} & q_{43} \\ q_{14} & q_{24} & q_{34} & q_{44} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{pmatrix} \\ = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Thus the affine transformation (1) is a motion if and only if its matrix Q satisfies condition (8).

Notice that the nonvanishing of the determinant Q follows automatically from the preceding condition; what is more, condition (8) implies that $(\det Q)^2 = 1$, or

$$\det Q = \pm 1 \quad (9)$$

§ 197. By means of formulas (6) it can easily be shown that the scalar product and the orthogonality of vectors are preserved under a motion in the Minkowski space. Indeed, assume that the vectors $x = \sum x_i e_i$, $y = \sum y_i e_i$ are mapped into the vectors $x' = \sum x'_i e_i$, $y' = \sum y'_i e_i$ (decomposed relative to the same basis) by a motion. Then

$$\begin{aligned} x'y' &= x'_1 y'_1 + x'_2 y'_2 + x'_3 y'_3 - x'_4 y'_4 = \sum_{i,k=1}^4 \sigma_{ik} x'_i y'_k \\ &= \sum_{i,k=1}^4 \sigma_{ik} \left(\sum_{\alpha=1}^4 q_{i\alpha} x_\alpha \right) \left(\sum_{\beta=1}^4 q_{k\beta} y_\beta \right) = \sum_{\alpha,\beta=1}^4 \left(\sum_{i,k=1}^4 \sigma_{ik} q_{i\alpha} q_{k\beta} \right) x_\alpha y_\beta \\ &= \sum_{\alpha,\beta=1}^4 \sigma_{\alpha\beta} x_\alpha y_\beta = x_1 y_1 + x_2 y_2 + x_3 y_3 - x_4 y_4 = xy \end{aligned}$$

Hence $x'y' = xy$. In particular, if $xy = 0$, then $x'y' = 0$, that is, the images of orthogonal vectors are orthogonal. This implies a very important result: in the Minkowski space every motion maps an orthonormal basis into an orthonormal basis.

§ 198. If under a motion a given point in space remains fixed, then the isotropic cone of the space with the vertex at this point also remains fixed (the points of the cone are displaced but remain on it). Indeed, assume that the origin O remains fixed. If M is any

point of the isotropic cone with vertex O , then $\rho(O, M) = 0$, and because a motion preserves the distance, it follows that for the image M' of M we have $\rho(O, M') = 0$; consequently, M' lies on the same isotropic cone. By analogous arguments one can show that under a motion the points lying in the interior of the isotropic cone remain there; it is not ruled out, however, that the sheets of the isotropic cone may interchange their position.

§ 199. The set of all motions in the Minkowski space constitutes a group, since the product of two motions is a motion and the inverse of a motion is also a motion.

These group properties evidently follow from the definition of motions. The group of motions in the Minkowski space is a subgroup of the affine group. From the group of all motions in the Minkowski space one can, in turn, single out a subgroup of motions which have a certain point fixed. A still narrower group is formed by motions under which every sheet of the isotropic cone retains its position.

§ 200. The transformation

$$X'_i = \sum_{k=1}^4 q_{ik} X_k + b_i, \quad \det q_{ik} \neq 0, \quad i = 1, 2, 3, 4$$

which represents a motion in the Minkowski space, is known as the *general Lorentz transformation*. The general Lorentz transformation is characterized by condition (8), § 196, on the matrix $Q = (q_{ik})$. The arbitrary quantities b_i do not play any significant role in the study of the general Lorentz transformations. Therefore usually the general Lorentz transformation means the homogeneous transformation, free of b_i 's,

$$X'_i = \sum_{k=1}^4 q_{ik} X_k, \quad i = 1, 2, 3, 4 \quad (1)$$

with the matrix Q satisfying the same condition as before. The set of all general Lorentz transformations constitutes a group, called the *general Lorentz group*.

If transformation (1) represents a motion in the Minkowski space under which every sheet of the isotropic cone retains its position, then such a transformation is simply called a *Lorentz transformation*. Besides condition (8), § 196, imposed on the matrix Q , these transformations are characterized by the fact that the point $(0, 0, 0, x_4)$, $x_4 > 0$, is mapped into the point (x'_1, x'_2, x'_3, x'_4) , $x'_4 > 0$. The Lorentz transformations constitute a group, the *Lorentz group*.

§ 201. The Lorentz transformations (or the general Lorentz transformations) may also be interpreted geometrically.

Let a coordinate system with origin O and orthonormal basis e_1, e_2, e_3, e_4 ($e_4^2 = -1$) be given; suppose further that a new coordinate system is introduced with the origin O' and the orthonormal

basis e'_1, e'_2, e'_3, e'_4 ($e'^2_3 = -1$). Then, if

$$e'_i = \sum_{k=1}^4 p_{ik} e_k \quad (1)$$

is the decomposition of the vectors of the new basis in terms of the old basis and if

$$X'_i = \sum_{k=1}^4 q_{ik} X_k + b_i \quad (2)$$

express new coordinates in terms of the old coordinates, the matrix $Q = (q_{ik})$ is the inverse of the transpose of the matrix $P = (p_{ik})$, i.e. $Q = (P^*)^{-1}$ (see § 180). Because both the new and old bases are orthonormal, it follows that

$$e_i e_k = \sigma_{ik}, \quad e'_i e'_k = \sigma_{ik}$$

in the notation of § 196. This yields

$$\begin{aligned} \sigma_{ik} = e'_i e'_k &= \sum_{\alpha=1}^4 p_{i\alpha} e_{\alpha} \sum_{\beta=1}^4 p_{k\beta} e_{\beta} = \\ &= \sum_{\alpha, \beta=1}^4 (e_{\alpha} e_{\beta}) p_{i\alpha} p_{k\beta} = \sum_{\alpha, \beta=1}^4 \sigma_{\alpha\beta} p_{i\alpha} p_{k\beta} \end{aligned} \quad (3)$$

If p^*_k are elements of P^* , that is, $p^*_{\beta k} = p_{k\beta}$, the last relation can be written as

$$\sum_{\alpha, \beta=1}^4 p_{i\alpha} \sigma_{\alpha\beta} p^*_{\beta k} = \sigma_{ik} \quad (i, k = 1, 2, 3, 4) \quad (4)$$

All these equations are equivalent to the single matrix equation

$$PIP^* = I \quad (5)$$

but, since P^* and Q are the inverse of each other, it follows that

$$P^*Q = E, \quad Q^*P = E$$

where E is the identity matrix. Premultiplication by Q^* and postmultiplication by Q of equation (5) yields

$$Q^*PIP^*Q = EIE = I = Q^*IQ \quad (6)$$

or

$$Q^*IQ = I$$

The last relation is exactly the same as (8), § 196. Hence any transformation of coordinates that maps one orthonormal system into another orthonormal system is the general Lorentz transformation (in general, inhomogeneous). Conversely, if the basis e_1, e_2, e_3, e_4 is orthonormalized and condition (6) is fulfilled, then we find, taking

note of the fact that (6) implies (5) then (4) and then (3), that $e'_i e'_k = \sigma_{ik}$, that is, the new basis is also orthonormalized. Hence any general Lorentz transformation can be treated as a transformation of orthonormalized coordinates. With this interpretation for general Lorentz transformations, among these transformations the simply Lorentz transformations are characterized by the fact that the basis vectors e_4, e'_4 lie on one sheet of the isotropic cone.

§ 202. We conclude the present section by pointing out a striking connection between the Lorentz group and the group of motions in Lobachevskian geometry. In order to make the treatment simple, we confine ourselves to the case of the three-dimensional Minkowski space with the isotropic cone

$$X_1^2 + X_2^2 - X_3^2 = 0 \quad (1)$$

If we cut this cone by the plane $X_3 = 1$, we obtain the circle

$$X_1^2 + X_2^2 = 1, \quad X_3 = 1 \quad (2)$$

which will be denoted by k . Suppose that the homogeneous Lorentz transformation

$$\left. \begin{aligned} X'_1 &= q_{11}X_1 + q_{12}X_2 + q_{13}X_3 \\ X'_2 &= q_{21}X_1 + q_{22}X_2 + q_{23}X_3 \\ X'_3 &= q_{31}X_1 + q_{32}X_2 + q_{33}X_3 \end{aligned} \right\} \quad (3)$$

is given; this transformation defines a motion in the Minkowski space that maps a point $M (X_1, X_2, X_3)$ into the point $M' (X'_1, X'_2, X'_3)$ and leaves the origin intact.

Let us associate a transformation of the plane $X_3 = 1$ with the given Lorentz transformation. Namely, if P is the point where the line OM meets the plane $X_3 = 1$ and if P' is the point where the line OM' meets the same plane $X_3 = 1$, then P' will be considered as the image of P . It is a simple matter to express this transformation in terms of coordinates.

Assume that $(x, y, 1)$ are the coordinates of P . Because O, P, M are collinear, we have

$$\frac{x}{X_1} = \frac{y}{X_2} = \frac{1}{X_3}$$

Accordingly,

$$x = \frac{X_1}{X_3}, \quad y = \frac{X_2}{X_3}$$

Similarly, if $(x', y', 1)$ are the coordinates of P' , then

$$x' = \frac{X'_1}{X'_3}, \quad y' = \frac{X'_2}{X'_3}$$

which in combination with formulas (3) yield

$$\left. \begin{aligned} x' &= \frac{q_{11}x + q_{12}y + q_{13}}{q_{31}x + q_{32}y + q_{33}} \\ y' &= \frac{q_{21}x + q_{22}y + q_{23}}{q_{31}x + q_{32}y + q_{33}} \end{aligned} \right\} \quad (4)$$

Since $\det q_{ik} \neq 0$, the mapping of the plane $X_3 = 1$ into itself, given by (4), is a projectivity (see § 112). If we take account of the fact that the Lorentz transformation given in space leaves the isotropic cone and its interior intact, we find that the transformation (4) of the plane $X_3 = 1$ leaves the circle k and its interior intact. Therefore transformation (4) is a non-Euclidean motion in the Lobachevskian metric defined on the plane $X_3 = 1$ in the interior of the absolute conic k (see § 170). We have demonstrated that every Lorentz transformation induces a non-Euclidean motion inside k on the plane $X_3 = 1$. Let us now show that given a non-Euclidean motion inside k , a unique Lorentz transformation inducing the given motion can be found.

Assume that a non-Euclidean motion is given by formulas of the form (4). Then the desired Lorentz transformation must be of the form

$$\left. \begin{aligned} X'_1 &= \lambda (q_{11}X_1 + q_{12}X_2 + q_{13}X_3) \\ X'_2 &= \lambda (q_{21}X_1 + q_{22}X_2 + q_{23}X_3) \\ X'_3 &= \lambda (q_{31}X_1 + q_{32}X_2 + q_{33}X_3) \end{aligned} \right\} \quad (5)$$

where the number $\lambda \neq 0$. For any $\lambda \neq 0$, formulas (5) define an affine transformation in the space. We shall prove that if λ is chosen suitably this affine transformation will become a Lorentz transformation.

In fact, transformation (4) leaves the circle k and its interior intact; this means that the affine transformation (5) does not affect the isotropic cone and its interior. Algebraically this means that (5) implies

$$X_1'^2 + X_2'^2 - X_3'^2 = \sigma (X_1^2 + X_2^2 - X_3^2) \quad (6)$$

where σ is proportional to λ^2 with a positive proportionality factor. Choose λ subject to $\sigma = 1$; then transformation (5) represents a motion in the Minkowski space. Notice also that $q_{33} \neq 0$, because otherwise the interior point $(0, 0, 1)$ of the isotropic cone will be mapped by (5) into the exterior point $(\lambda q_{13}, \lambda q_{23}, 0)$ which is impossible. Therefore we can choose the sign of λ such that $\lambda q_{33} > 0$. Under this condition, transformation (5) leaves every sheet of the isotropic cone intact and hence is a Lorentz transformation. It is clear that the choice of λ is unique.

Thus the three-dimensional homogeneous Lorentz transformations are in a one-to-one correspondence with motions in two-dimensional

Lobachevskian geometry. It can be easily verified that to the product of two Lorentz transformations there is assigned the product of the corresponding non-Euclidean motions. Hence the three-dimensional homogeneous Lorentz group and the group of motions in two-dimensional Lobachevskian geometry are isomorphic.

One can analogously show that the four-dimensional homogeneous Lorentz group and the group of motions in three-dimensional Lobachevskian geometry are isomorphic.

7.3. The space of events of the special relativity theory

§ 203. Consider an event M . For us the nature of this event is immaterial and we are only concerned with the place and the time of its occurrence. Suppose that M occurs in such a small part of space and for such a small interval of time that it can be treated as occurring at a definite point for a moment. Under the above assumptions event M will be called elementary. The place of occurrence of any elementary event is determined relative to a preassigned material body and time is determined according to a definite clock. For example, one can determine the place of every event relative to the earth and time according to the clock of the Pulkovo observatory*.

Suppose that a material body T is chosen relative to which the place of occurrence of any elementary event is determined; suppose that three mutually perpendicular Cartesian axes are associated with T and a scale is given relative to which the place of event M is characterized by the coordinates x, y, z (assuming the geometric properties of the real space to be Euclidean); suppose, finally, that a clock is given according to which the moment of occurrence of event M is characterized by a number t (assuming t to be equal to the number of time units from a given moment). The set composed of T , the scale, the axes, the watch and the initial moment is called a *frame of reference* and the quantities x, y, z, t are called the *coordinates* of event M in the given reference frame.

If the reference frame chosen is changed, then the same event M will have, in general, other coordinates x', y', z', t' in a new reference frame. If the same body T is taken while the axes, scale, time unit and the initial moment connected with T are changed, then the change in reference frame and the corresponding transformation of the coordinates of the events are called *trivial*. In contrast, the new reference frame and the corresponding transformation of coordinates of the events are *nontrivial* if one takes another body T' , instead of T , that moves relative to T .

* The Central Observatory of the USSR Academy of Sciences, located near Leningrad.

Of vital interest for physics is the question as to how the coordinates of events change when the change in the reference frame is non-trivial. Of course, such a question is sensible only for sufficiently specific classes of reference frames. This question is discussed below for the inertial frames of reference.

§ 204. A reference frame S is called *inertial* if every free particle moves uniformly in a line with respect to the frame S . When we speak of a free particle we have in mind a body of small dimensions so distant from other bodies that their action on this body can be neglected.

Let S and S' be two inertial reference frames and M any event. Our aim is to obtain or characterize formulas expressing the coordinates (x', y', z', t') of M in the frame S' in terms of the coordinates (x, y, z, t) of the same event in frame S .

Let us examine the solution of this problem from the point of view of classical physics. First of all, classical physics assumes that it is possible to synchronize all the clocks and hence that it is possible to establish one time unit for all inertial reference frames; thus $t = t'$. Next, it is assumed that it is possible to establish one scale for measurement of the lengths of segments of all the coordinate axes of the frames S and S' . These assumptions and the law of addition of velocities of classical kinematics imply that under a special choice of the coordinate axes in frames S and S' the coordinates of any event M change according to the formulas

$$x' = x - vt, \quad y' = y, \quad z' = z, \quad t' = t \quad (1)$$

when the frame S is changed to S' (the coordinate axes are chosen such that the axis $O'x'$ lies on the axis Ox and the axes $O'y'$, $O'z'$ remain parallel to the axes Oy , Oz ; v is the velocity of S' with respect to S).

Thus the assumptions of classical physics enable us to obtain the desired formulas easily and these formulas have a very simple form.

It should be noted, however, that the possibility of the general synchronization of clocks is by no means as obvious as it might appear at first. One could synchronize clocks in all inertial frames of reference if there existed signals that could propagate instantly. It would have been enough to put the clocks in one inertial frame in a definite phase and then immediately send the signal to other frames and there put the clocks in the same phase at the time of receiving the signal; then the clocks could be unified by sending a second signal after a definite interval of time. In such a case all the inertial frames would enjoy equal privilege in the sense that the despatch of a signal, from any one frame and its receipt in any other reference frame would be done under the same phases of the clocks of these frames. But there are no signals in nature that get pro-

pagated instantly. If one uses light signals, then by the method just described only an approximate synchronization of clocks can be made on inertial systems the velocity of which relative to other frames is small in comparison to the velocity of light.

Similarly, there can be no objection in the approximate sense regarding the other two assumptions which were referred to above (the possibility of unification of scales and the classical law of addition of velocities). Accordingly, formulas (1) are true only approximately, provided v is small in comparison to the velocity of light.

But formulas (1) contradict the experimental results of the modern physics of large velocities. The prevailing situation is as follows. It was discovered a long time ago that the laws of mechanics are the same in all inertial frames of reference. This fact is not contradicted by formulas (1) if the laws of classical mechanics are properly interpreted, since the equations (of classical mechanics) are invariant under the transformation given by formulas (1). At the same time, (1) imply that the laws of electrodynamics must depend on the inertial frame chosen, because the equations of electrodynamics are not invariant under transformation (1). First of all, the velocity of light must be different for different inertial frames; namely, if in frame S light is propagated along the x -axis with velocity c , then according to formulas (1) the velocity of light in S' must be $c - v$. However, this phenomenon was not confirmed by necessary experiments. In light of this, in physics one accepts the postulate stating that the velocity of light is independent of the choice of inertial frame of reference. This is the starting point for the special theory of relativity discovered by Lorentz, Poincaré, Minkowski and especially Einstein, according to which not only the laws of mechanics but also the laws of electrodynamics are the same in all inertial frames of reference. The initial assumptions of classical physics that led to formulas (1) are replaced in the theory of relativity by more precise assumptions which agree with experimental physics of large velocities. This leads to the replacement of formulas (1) with more precise formulas, which will be deduced in the following subsections. In the process, we shall have to make use of geometric concepts developed in the preceding two sections.

§ 205. Let S be an inertial frame of reference, M any elementary event, and t, x, y, z the coordinates of M in S (here and henceforth time t is taken as the first coordinate to facilitate the writing of certain formulas).

Denote by $\mathfrak{A}$ a four-dimensional affine space in which the origin O and the basis a_1, a_2, a_3, a_4 for the affine coordinate system are chosen in an arbitrary manner. With M we associate that point of $\mathfrak{A}$ which is determined by the coordinates t, x, y, z relative to the chosen origin and basis; we shall say that this point represents an event M

in space $\mathfrak{M}$. The point representing the event will be denoted by the same letter as the event itself.

The four-dimensional affine space $\mathfrak{A}$ whose points represent all possible elementary events is called the *space of events*. Notice that the events which take place during a time interval $t_1 \leq t \leq t_2$ at a point of physical space which is fixed relative to the axes of S and has coordinates x_0, y_0, z_0 are represented in $\mathfrak{A}$ by the segment $t_1 \leq t \leq t_2, x = x_0, y = y_0, z = z_0$, which is evidently parallel to the vector a_1 . On this account, the coordinate axis which is directed in the space of events along the basis vector a_1 is called the *time axis*.

§ 206. We now take the first step towards solving the problem of transformation of coordinates of events when we move one inertial frame of reference to another. Suppose that another inertial reference frame S' , besides S , is taken, and let t, x, y, z be the coordinates of any event M in S and t', x', y', z' those of the same event in S' . Then t', x', y', z' are definite functions of t, x, y, z :

$$\left. \begin{aligned} t' &= f(t, x, y, z) \\ x' &= \varphi(t, x, y, z) \\ y' &= \psi(t, x, y, z) \\ z' &= \chi(t, x, y, z) \end{aligned} \right\} \quad (1)$$

Assume that (a) f, φ, ψ, χ are defined for all values of t, x, y, z and (b) t, x, y, z are uniquely obtained from (1) by prescribing any value to t', x', y', z' . These assumptions amount to assuming that the events may take place relative to frames S and S' anywhere and at any time; these assumptions also imply that the frames S, S' are inertial all the time. We shall show that

Formulas (1) are linear, that is, are of the form

$$\left. \begin{aligned} t' &= c_{11}t + c_{12}x + c_{13}y + c_{14}z + d_1 \\ x' &= c_{21}t + c_{22}x + c_{23}y + c_{24}z + d_2 \\ y' &= c_{31}t + c_{32}x + c_{33}y + c_{34}z + d_3 \\ z' &= c_{41}t + c_{42}x + c_{43}y + c_{44}z + d_4 \end{aligned} \right\} \quad (2)$$

and, moreover, that $\det c_{ih} \neq 0$.

Proof. To prove this theorem, we shall have to make one more physical assumption. Namely, we shall assume that at any instant of time through any place in the physical space a free particle can fly in any direction with any velocity known in experimental physics (we do not assume, however, that the particle can fly with any *arbitrary* velocity, because arbitrarily large velocities have not been observed and such an assumption would not be justified; what is more, the results below will confirm that such an assumption is inadmissible). The velocity of the particle will mean its velocity

relative to S . Denote by C a positive constant such that any velocity less than C can be achieved.

Suppose that a free particle flies in space. Because S is an inertial frame, the motion of the particle in question with respect to S is uniform and is along a straight line. Therefore the equations of motion of this particle must be of the form

$$x - x_0 = l(t - t_0), \quad y - y_0 = m(t - t_0), \quad z - z_0 = n(t - t_0) \quad (3)$$

where l, m, n are the velocity components of the flying particle, and (x_0, y_0, z_0) is the place where the particle is at the instant $t = t_0$. In view of the assumptions made in the beginning of the proof, the quantities t_0, x_0, y_0, z_0 may be arbitrary but l, m, n must be such that

$$l^2 + m^2 + n^2 < C^2 \quad (4)$$

The fact that the flying particle is at (x, y, z) at instant t is an event, which is represented by the point (t, x, y, z) in the space of events $\mathfrak{A}$. The entire process of motion of the flying particle is represented by a certain line in the space of events, because t, x, y, z are connected by three independent first-order equations (3) (see § 185); we call this line b . Notice further that relations (3) and (4) imply

$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 - C^2(t - t_0)^2 < 0 \quad (5)$$

which defines the interior of a certain real nondegenerate quadric cone with vertex (t_0, x_0, y_0, z_0) (see § 186); call this cone K_0 . Because inequality (5) is a consequence of (3) and (4), the line b through the vertex of K_0 lies in its interior. It also follows from the assumptions made that every line in the space of events which passes inside the cone K_0 through its vertex may represent the process of motion of a free particle flying past (x_0, y_0, z_0) at instant t_0 .

Let us now examine equations (1). By virtue of these equations, with every point $M(t, x, y, z)$ in the space of events there is associated a point $M'(t', x', y', z')$, that is, a mapping $M' = f(M)$ is defined. In view of the conditions imposed on equations (1), this mapping is a one-to-one mapping of the space of events into itself. If we now take account of the fact that the reference frame S' is also inertial, we find, letting t, x, y, z denote the current coordinates in equations (3), that t', x', y', z' satisfy analogous equations, although with different parameters (because equations (3) define a motion of a free particle and this motion in S' is uniform and rectilinear). This implies that if the points M lie on a line b in the space of events, the corresponding points $M' = f(M)$ also lie on some line b' . Summing up the above discussion, we have the following: (i) for any point $M_0(t_0, x_0, y_0, z_0)$ in the event space there is defined a cone K_0 with vertex M_0 , and (ii) if the line b passes through M_0 in the interior of K_0 , then the images under the mapping $M' = f(M)$ of all points of b lie on the line b' . Now we shall show that whatever the line b ,

the images of all its points also lie on one line, that is, the mapping $M' = f(M)$ is a collineation.

Take three distinct points M_1, M_2, M_3 on b and let K_1, K_2, K_3 be the cones defined for M_1, M_2, M_3 in the same way as K_0 was defined for the point M_0 . It is now natural to assume that b does not pass through the interiors of K_i . Through M_1 we pass a line b_1 inside K_1 . The lines b and b_1 determine a (two-dimensional) plane β containing these lines. Through M_2 and M_3 we pass lines b_2^* and b_3^* parallel to b_1 ; these lines lie in β and inside the respective cones K_2 and K_3 . Since the left side of the inequality (5) is continuous, it follows that if the coordinates of the direction vectors of b_2^* and b_3^* undergo a small change, the new direction vectors define lines also lying inside the cones K_2 and K_3 . Therefore there are lines b_2 and b_3 which pass through M_2 and M_3 in β and inside the cones K_2 and K_3 and are situated so that all three lines b_1, b_2, b_3 intersect pairwise in three distinct points P, Q, R of β .

Let P', Q', R' be images of P, Q, R under the mapping $M' = f(M)$. The points P', Q', R' are distinct because the given mapping is one-to-one. If R' lies on the line $P'Q'$, then $M'_i = f(M_i)$ ($i = 1, 2, 3$) also lie on the same line, and nothing more is left to be proved. Assume that P', Q', R' are noncollinear. Then these points determine a plane β' . Since the lines b_1, b_2, b_3 pass inside the cones K_1, K_2, K_3 , the images of the points of these lines lie on three lines b'_1, b'_2, b'_3 . The lines b'_1, b'_2, b'_3 intersect pairwise at P', Q', R' and hence lie in β' (see § 182); simultaneously, β' also contains the points $M'_i = f(M_i)$ ($i = 1, 2, 3$). Through M_1 one may pass a line c_1 inside K_1 that does not lie in β . Repeating with respect to c_1 the same construction as was done with respect to b_1 , we obtain a plane γ' , distinct from β' , containing M'_1, M'_2, M'_3 . Since β' and γ' are distinct, all their common points are collinear, implying that the points M'_1, M'_2, M'_3 are collinear and hence proving that $M' = f(M)$ is a collineation. But if $M' = f(M)$ is a collineation, then, according to § 187, it is represented in affine coordinates by linear formulas with a non-vanishing determinant. ■

Remark. The above proof continues to hold if the reference frame S' is assumed to be inertial with respect to the frame S , that is, if it is only assumed that any rectilinear uniform motion of the particle (with an admissible velocity) relative to S is also rectilinear and uniform relative to S' . The above proof did not make use of the mutuality of such a relationship between S and S' .

We now know that the desired formulas for the transformation of coordinates of events are of linear form. Below we shall find the coefficients in these formulas by assuming the postulate regarding the independence of the velocity of light from the inertial reference-frame chosen. But first we must add something to the concept of the event space described in § 205.

§ 207. In defining the space of events $\mathfrak{A}$, we started by taking an inertial reference frame S in the physical space. Then in the affine space $\mathfrak{A}$ we chose the origin O and the basis a_1, a_2, a_3, a_4 of the affine coordinate system; the point M representing the event is determined in the chosen system by the coordinates (t, x, y, z) relative to S . If S' is any other inertial reference frame in which the same event has coordinates (t', x', y', z') , then the new coordinates are expressed in terms of t, x, y, z by linear formulas (2), § 206, with a non-vanishing determinant. But, in view of § 180, any linear transformation of four variables with a nonvanishing determinant can be treated as the transformation of affine coordinates in the four-dimensional affine space. Thus there is a new affine coordinate system in $\mathfrak{A}$ defined by an origin O' and a basis a'_1, a'_2, a'_3, a'_4 in which the point M representing the event in question has the coordinates (t', x', y', z') . Hence with every inertial reference frame in the physical space there is associated a definite affine coordinate system in the event space.

§ 208. Let M_0 and M be two events which are observed on an inertial reference frame S . Assume that M_0 occurs at a given place (x_0, y_0, z_0) of the physical space at a given instant of time t_0 and the place (x, y, z) of event M is arbitrary; let t denote the time of occurrence of M . Suppose that if $t_0 < t$, the light signal sent from the place of event M_0 at instant t_0 is received at the place of event M exactly at instant t , and if $t < t_0$, the light signal informing the event M is received at the place of event M_0 exactly at instant t_0 . Then

$$-c^2(t-t_0)^2 + (x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 = 0 \quad (1)$$

where c is the velocity of light, because the path described by the signal equals its velocity multiplied by the corresponding interval of time. In space $\mathfrak{A}$, the events M_0 and M are represented by two points, also denoted by the same letters M_0 and M , in which M_0 is a fixed point and M the variable point. The set of all points M in $\mathfrak{A}$ determined by equation (1) is a cone with vertex M_0 [(see § 186); this cone is known as the *light cone* of the event space at the point M_0 . Equation (1) defines the light cone in the affine system O, a_1, a_2, a_3, a_4 which is associated with the inertial reference frame S .

The interior of the light cone is characterized by the inequality

$$-c^2(t-t_0)^2 + (x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2 < 0$$

This inequality in combination with the condition $t > t_0$ defines one sheet of the light cone which we shall call the *upper sheet*. To the upper sheet there correspond those events M to which the light signal informing about the event M_0 reaches before these events take place. The second sheet of the light cone corresponding to the condition $t < t_0$ is called the *lower sheet*; to this sheet there correspond events M which precede M_0 and the light signal informing about the

event M also reaches the place of occurrence of event M_0 before this event.

The exterior of the light cone corresponds to those events M which cannot be connected by information with the event M_0 even by means of light signals (this will be the case if, for example, M_0 and M take place simultaneously at different places of the system S , that is, are such that $t = t_0$).

§ 209. We are now in a position to discuss how to obtain coefficients in formulas for transformation of coordinates of events when the inertial reference frame is altered. Let us change to the affine system O' , a'_1 , a'_2 , a'_3 , a'_4 corresponding to the reference frame S' . In the new system the points M_0 and M have new coordinates but the light cone is defined by the equation

$$-c^2(t' - t'_0)^2 + (x' - x'_0)^2 + (y' - y'_0)^2 + (z' - z'_0)^2 = 0 \quad (2)$$

analogous to the preceding equation, with the same value of c . This is a consequence of the postulate regarding the independence of the velocity of light from the inertial reference frame chosen.

Transforming equation (2) to old coordinates t , x , y , z by using formulas (2), § 206, we have

$$\begin{aligned} -c^2(t' - t'_0)^2 + (x' - x'_0)^2 + (y' - y'_0)^2 + (z' - z'_0)^2 \\ = A(t - t_0)^2 + B(t - t_0)(x - x_0) + \dots + K(z - z_0)^2 \end{aligned} \quad (3)$$

Here the right hand side is a quadratic form in the differences $t - t_0$, $x - x_0$, $y - y_0$, $z - z_0$ whose coefficients A , B , $\dots$, K are certain expressions in terms of the coefficients c_{ih} of formulas (2), § 206, and c .

If the right hand side of equation (3) is equated to zero, the equation of the light cone in old coordinates will be obtained. But in old coordinates this cone is defined by equation (1); accordingly, the coefficients of the right side of equation (3) must be proportional to the coefficients of equation (1).

Therefore equation (3) must be of the form

$$\begin{aligned} -c^2(t' - t'_0)^2 + (x' - x'_0)^2 + (y' - y'_0)^2 + (z' - z'_0)^2 \\ = H \{ -c^2(t - t_0)^2 + (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \} \end{aligned} \quad (4)$$

where H is a certain expression composed of the coefficients c_{ih} present in formulas (2), § 206; moreover $H > 0$ (as a consequence of the law of inertia of quadratic forms). Assume now that in one of the two reference frames S and S' the linear scale and the unit of time are changed by the same quantity; in that case all the differences $t - t_0$, $x - x_0$, $y - y_0$, $z - z_0$ are multiplied by the same number (simultaneously, the coefficients c_{ih} in formulas (2), § 206, undergo proportional change). Therefore, under compatible linear scales and units of time the relation $H = 1$ can be made to hold in the frames

S and S' . Henceforth it is assumed that such a compatibility has been achieved in every case. Then for any pair of points M_0 and M we must have

$$\begin{aligned} -c^2(t' - t'_0)^2 + (x' - x'_0)^2 + (y' - y'_0)^2 + (z' - z'_0)^2 \\ = -c^2(t - t_0)^2 + (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \end{aligned} \quad (5)$$

as a consequence of formulas (2), § 206. Thus in order that formulas (2), § 206, express the transformation of coordinates of events under a change from one inertial reference frame into another, the coefficients must be chosen so that relation (5) holds identically. These are the necessary conditions on the coefficients of the desired formulas. What is more, these conditions happen to be sufficient as well; namely, as the following analysis will reveal, under these conditions the arbitrary choice of the coefficients implies the arbitrary choice of the possible inertial frames.

We need not carry out any computations in order to obtain the conditions about which we are talking. The desired result has been essentially obtained in the preceding section where we examined the Minkowski space; what we have to do is to be able to extract this condition from there. To this end, we introduce the Minkowski metric in the space of events.

With any pair of points M_0, M of the event space we associate the number

$$\rho(M_0, M) = \sqrt{-c^2(t - t_0)^2 + (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2} \quad (6)$$

which will be called the *distance* between the points (or events) M_0, M . Simultaneously, we define the scalar product of any pair of vectors M_0M, M_0N by taking it equal to the bilinear form in the coordinates of vectors M_0M, M_0N ; the coefficients of this bilinear form are naturally taken to be the same as those of the quadratic form present under the square root in formula (6) (see § 191).

Formula (6) expresses the distance between the events in coordinates relative to a definite inertial reference frame S . But as shown by relation (5), the choice of the inertial reference frame here is immaterial. In other words, the Minkowski metric introduced in the space of events is invariant under the change from one inertial reference frame to another.

The distance between two events M_0 and M may be imaginary, zero or positive. Namely, $\rho(M_0, M)$ is imaginary if M_0 and M can be connected by information with the help of a signal whose velocity $< c$; $\rho(M_0, M) = 0$ if the information is possible only by using the light signal; $\rho(M_0, M) > 0$ if even the light signal sending information about one event cannot alert the other. These three classes of events M correspond respectively to the interior of the light cone with vertex M_0 , the surface of the cone itself, and the exterior of this cone (see § 208). We note in passing that the light

cone of the space of events is nothing but the isotropic cone of the Minkowski metric introduced in this space.

Formula (6) implies that the norm of the first basis vector of the selected affine coordinate system is ci and the norms of the remaining basis vectors are unity. Replacing the first basis vector by an imaginary-unit vector in the same direction and leaving the remaining basis vectors as they are, we shall now denote the basis vectors by e_1, e_2, e_3, e_4 (in § 207 these vectors were denoted by a_1, a_2, a_3, a_4). Thus

$$e_1 = (1/c) a_1, \quad e_2 = a_2, \quad e_3 = a_3, \quad e_4 = a_4$$

and the first coordinate of the event is now ct , instead of t . We introduce new symbols for the coordinates of the event:

$$ct = X_1, \quad x = X_2, \quad y = X_3, \quad z = X_4$$

The coordinates X_i will be called the normalized coordinates of the event. Formula (6) yields

$$\rho^2(M_0, M) = -(X_1 - X_1^0)^2 + (X_2 - X_2^0)^2 + (X_3 - X_3^0)^2 + (X_4 - X_4^0)^2 \quad (7)$$

where X_h^0 and X_h are normalized coordinates of the events M_0 and M . In accordance with formula (7), the scalar product of two vectors x, y with coordinates x_h, y_h is given by

$$xy = -x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 \quad (8)$$

Since formulas (7) and (8) have canonical forms, the basis e_1, e_2, e_3, e_4 is orthonormalized and $e_1^2 = -1$ (see § 193). Suppose that in the physical space we change from the inertial frame S to the inertial frame S' . Then the coordinates of the events are transformed according to linear formulas; we write them for normalized coordinates in the compact form

$$X'_i = \sum_{k=1}^4 q_{ik} X_k + b_i, \quad i = 1, 2, 3, 4 \quad (9)$$

This transformation amounts to a change in the space of events from the affine system O, e_1, e_2, e_3, e_4 associated with S to the affine system $O', e'_1, e'_2, e'_3, e'_4$ associated with S' . Because the metric in the space of events is invariant, the basis e'_1, e'_2, e'_3, e'_4 is also orthonormalized and $e'^2_1 = -1$. Hence, if in the physical space one inertial reference frame is replaced with another, then the coordinates of any event (X_1, X_2, X_3, X_4) are transformed by formulas (9), which coincide with the formulas expressing the transformation of orthonormalized coordinates in the four-dimensional Minkowski space.

Thus we have basically solved the problem posed, because the question of transforming orthonormalized coordinates was studied in the preceding section (see § 201), only one detail remains to be examined which will be done now.

§ 210. To make the presentation simple, notice that by means of a trivial transformation of the inertial reference frames S and S' formulas (9) may be made homogeneous. Indeed, let us recall that the inertial reference frame is a system composed of a material body T , three Cartesian axes which are closely connected with T , a scale, clock, and the initial moment of reference. Moving parallel the Cartesian axes of one of the frames S and S' , we can make them located in such a way that at a certain instant the origin of the axes of frame S coincides with that of frame S' ; we take the moment of observation of this event in S and S' as the initial moment of reading in both of these systems. Then to the zeroes values of t, x, y, z there correspond zero values of t', x', y', z' . Hence the free terms b_i in formulas (9) become zero and we obtain

$$X'_i = \sum_{h=1}^4 q_{ih} X_h \quad (1)$$

In the space of events, formulas (1) correspond to a change in the basis of orthonormalized coordinates without changing the origin. Consider the light cone with origin as the vertex. Because the first vectors e_1, e'_1 of the old and new bases are imaginary-unit vectors, both of them lie in the interior of the light cone. Since the upper sheet of the light cone corresponds to the future events relative to the zero moment of the time reading in S , it follows that e_1 is indeed directed towards the upper sheet; but then e'_1 also must be directed towards the same sheet of the light cone. This implies that the transformation (1) of normalized coordinates of events corresponds in the space of events to the shift to the new orthonormalized basis under the condition that the new basis vector e'_1 lies in the same sheet of the light cone as the old basis vector e_1 .

In view of § 201, this assertion may be formulated as follows: every transformation of normalized inertial coordinates of events is the four-dimensional Lorentz transformation (here the Lorentz transformation is to be understood in the strict sense of the word).

§ 211. We can now immediately obtain the conditions that must be satisfied by the coefficients q_{ih} if formulas (1), § 210, are to express a transformation of normalized inertial coordinates of events; it is enough to refer to the matrix relation (6), § 201 (see also § 196). It should only be borne in mind that now e_1 , and not e_4 as in §§ 196-201, is the imaginary-unit vector. Therefore in the given case

$$I = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Thus, if formulas (1), § 210, express the transformation of normalized inertial coordinates, the matrix $Q = (q_{ih})$ satisfies the con-

dition

$$Q^*IQ = I, \quad q_{11} > 0 \quad (2)$$

(the condition $q_{11} > 0$ implies that e_1 and e'_1 lie in one sheet of the light cone).

§ 212. The matrix Q can be put in a very simple form if suitable trivial transformations of the inertial frames in question are performed.

Suppose that in the inertial frame S we change to new Cartesian axes but the time reading and the scale remain the same. Then the coordinates of events undergo a trivial transformation with the matrix

$$Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & q_{22} & q_{23} & q_{24} \\ 0 & q_{32} & q_{33} & q_{34} \\ 0 & q_{42} & q_{43} & q_{44} \end{pmatrix}$$

because $t' = t$ and x', y', z' are expressed in terms of x, y, z only. The matrix

$$Q_1 = \begin{pmatrix} q_{22} & q_{23} & q_{24} \\ q_{32} & q_{33} & q_{34} \\ q_{42} & q_{43} & q_{44} \end{pmatrix}$$

is the usual three-dimensional orthogonal matrix, well known in analytic geometry. The transformation of coordinates with the given matrix Q amounts to a change from the basis e_i to the basis e'_i in the space of events, and if

$$e'_i = \sum_{h=1}^4 p_{ih} e_h$$

then the matrix $P = (p_{ih}) = (Q^*)^{-1}$ (see § 180). Hence

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & p_{22} & p_{23} & p_{24} \\ 0 & p_{32} & p_{33} & p_{34} \\ 0 & p_{42} & p_{43} & p_{44} \end{pmatrix}$$

implying that $e'_1 = e_1$ and the vectors e'_2, e'_3, e'_4 are expressed in terms of the vectors e_2, e_3, e_4 only by formulas whose coefficients constitute the matrix

$$P_1 = \begin{pmatrix} p_{22} & p_{23} & p_{24} \\ p_{32} & p_{33} & p_{34} \\ p_{42} & p_{43} & p_{44} \end{pmatrix}$$

Evidently, $P_1 = (Q_1^*)^{-1}$, and, accordingly, P_1 is also a three-dimensional orthogonal matrix. Consequently, the vectors e'_2, e'_3, e'_4 lie in the hyperplane determined by the vectors e_2, e_3, e_4 and they may be obtained by means of the usual Euclidean motion of the triple of the vectors e_2, e_3, e_4 as a whole. One should note here that on the hyperplane $e_2e_3e_4$ three-dimensional Euclidean geometry is valid (see § 195). Conversely, if in the space of events $e'_1 = e_1$ and the vectors e'_2, e'_3, e'_4 are obtained from the triple of vectors e_2, e_3, e_4 by a Euclidean motion in the hyperplane $e_2e_3e_4$, then such a transformation corresponds to the trivial change to new Cartesian axes in the system S .

Let now S and S' be any two inertial reference frames (compatible in the sense of § 210) and e_i, e'_i the corresponding basis in the space of events (origins O and O' are assumed to coincide). If $e'_1 = e_1$, then S' is obtained from S by a trivial transformation, which is clearly of little interest. We shall assume that $e'_1 \neq e_1$ (the case $e'_1 = -e_1$ is impossible, see § 210). Let α denote the hyperplane determined by the vectors e_2, e_3, e_4 and α' the one determined by the vectors e'_2, e'_3, e'_4 . These hyperplanes have a common point O and are distinct. Therefore, in view of § 182, the hyperplanes α and α' intersect along a two-dimensional plane β . We leave the vectors e_1 intact and change the triple of vectors e_2, e_3, e_4 by a Euclidean motion on α in a new position such that e_2, e_3 lie in β . Analogously, leaving e'_1 intact, we change the triple of vectors e'_2, e'_3, e'_4 by a Euclidean motion on α' in a position where e'_2 and e'_3 also lie in β and, moreover, coincide with e_2 and e_3 . The notation of vectors is kept unchanged. To every change in the basis e_i and e'_i there corresponds a trivial transformation of frames S, S' of the physical space.

We now have

$$\begin{aligned} e'_1 &= p_{11}e_1 + p_{12}e_2 + p_{13}e_3 + p_{14}e_4 \\ e'_2 &= p_{21}e_1 + p_{22}e_2 + p_{23}e_3 + p_{24}e_4 \\ e'_3 &= \theta + \theta + e_3 + \theta \\ e'_4 &= \theta + \theta + \theta + e_4 \end{aligned}$$

Multiplying scalarly the two sides of the first equation by the corresponding sides of the third equation and noting that $e'_1e'_3 = 0$, $e_1e_3 = 0$, $e_2e_3 = 0$, $e_4e_3 = 0$, $e_3^2 = 1$, we see that $p_{13} = 0$. Similarly, we can demonstrate that $p_{14} = 0$, $p_{23} = 0$, $p_{24} = 0$. Accordingly, the matrix P is of the form

$$P = \begin{pmatrix} p_{11} & p_{12} & 0 & 0 \\ p_{21} & p_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

whence

$$Q = (P^*)^{-1} = \begin{pmatrix} q_{11} & q_{12} & 0 & 0 \\ q_{21} & q_{22} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1)$$

Thus by a trivial change of the inertial frames S , S' , the formulas expressing the normalized coordinates of events can always be put in the form

$$\left. \begin{aligned} X'_1 &= q_{11}X_1 + q_{12}X_2 \\ X'_2 &= q_{21}X_1 + q_{22}X_2 \\ X'_3 &= X_3 \\ X'_4 &= X_4 \end{aligned} \right\} \quad (2)$$

§ 213. To determine the remaining coefficients one could use conditions (2), § 211. However, for the special frames S and S' the desired transformation has such a simple form that we prefer to obtain the final result directly, starting with identity (5), § 209, which expresses the invariance of the metric of the space of events.

Let us go back to the physical notation of the coordinates of events, the preceding equations can then be written in the form

$$t' = At + Bx, \quad x' = Dt + Ex, \quad y' = y, \quad z' = z \quad (1)$$

For the given case, identity (5), § 209, assumes the form

$$-c^2 t'^2 + x'^2 = -c^2 t^2 + x^2$$

Substitution of expressions (1) into the left side of the last identity and comparison of coefficients of the resulting quadratic form with those of the right side yield

$$-c^2 A^2 + D^2 = -c^2 \quad (2)$$

$$-c^2 AB + DE = 0 \quad (3)$$

$$-c^2 B^2 + E^2 = 1 \quad (4)$$

Now note that from physical considerations, described in § 210 (see also § 211), one must have $A > 0$. The same considerations also imply that $E \neq 0$, and, moreover, one can assume $E > 0$ (by choosing a suitable direction on the x -axis). In view of (3), we have

$$\frac{D}{cA} = \frac{cB}{E}$$

Denoting each of these ratios by $-\beta$, we have $D = -\beta cA$, $cB = -\beta E$. These together with (2) and (4) yield

$$A = \frac{1}{\sqrt{1-\beta^2}}, \quad E = \frac{1}{\sqrt{1-\beta^2}}$$

from which B and D can be found. Thus

$$t' = \frac{t - \frac{\beta}{c}x}{\sqrt{1-\beta^2}}, \quad x' = \frac{-\beta ct + x}{\sqrt{1-\beta^2}}, \quad y' = y, \quad z' = z \quad (5)$$

We made use of all the algebraic conditions. Accordingly, the parameter β is arbitrary, or, more precisely, from the point of view of mathematics it must satisfy only the inequality

$$1 - \beta^2 > 0.$$

This parameter has a simple physical interpretation. To discover this, consider a point M in the physical space which is fixed in the frame S' ; the coordinates x' , y' , z' of this point are constants. The point M moves with respect to the frame S ; differentiation of the last three equations in (5) gives the velocity of M in S :

$$\frac{dx}{dt} = \beta c, \quad \frac{dy}{dt} = 0, \quad \frac{dz}{dt} = 0$$

Thus all the points which are fixed in S' move with respect to S with one and the same velocity $(\beta c, 0, 0)$ in the direction of the x -axis. This velocity, common to all the points of S' , is called the velocity of the motion of frame S' relative to S . Denoting by v the component of this velocity in the direction of the x -axis, we have $\beta = v/c$. Inequality (6) yields

$$v^2 < c^2 \quad (7)$$

If the physical space does not impose any other restrictions on the velocity with which one inertial frame of reference moves relative to another, then from the point of view of physics the parameter β is restricted only by inequality (6). Hence in this case there are no other conditions, apart from relations (2), § 211, imposed on the coefficients q_{ik} of transformation (1), § 210, and the problem posed before us has been completely solved.

In other words, every transformation of normalized inertial coordinates is a four-dimensional Lorentz transformation; every four-dimensional Lorentz transformation can be treated as a transformation of normalized inertial coordinates.

At the same time, one can say that the transformations of normalized inertial coordinates constitute a group which is isomorphic to the four-dimensional Lorentz group. The homogeneous transformations of normalized inertial coordinates constitute a group which is isomorphic to the four-dimensional homogeneous Lorentz group and is hence isomorphic to the group of non-Euclidean motions of three-dimensional Lobachevskian geometry (see § 202).

§ 214. We note a few results which follow from our discussion and computations.

(1) Inequality (7), § 213, implies that the velocity with which one inertial reference frame moves relative to another can only be less than the velocity of light.

(2) Substitution of $\beta = v/c$ into formulas (5) yields

$$t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad x' = \frac{-vt + x}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad y' = y, \quad z' = z \quad (1)$$

If v is small in comparison to the velocity of light c , then

$$t' \approx t, \quad x' \approx x - vt, \quad y' = y, \quad z' = z$$

and thus the approximate formulas of classical physics are obtained (see § 204).

(3) If two events (t_1, x_1) , (t_2, x_2) occur at different points of the x -axis in frame S but simultaneously ($t_1 = t_2$), the first formula in (1) implies that $t'_1 \neq t'_2$, because $x_1 \neq x_2$. Thus the events taking place simultaneously in S do not occur simultaneously in S' . Hence the general synchronization of clocks assumed in classical physics is impossible (in this connection see § 204).

(4) Suppose that a pole of length $\Delta x' = x'_2 - x'_1$ lies in the equilibrium position on the x' -axis in S' , and suppose that in S in which this pole moves it is measured at a definite moment t . The second equation in (1) yields

$$\Delta x = x_2 - x_1 = \Delta x' \sqrt{1 - \frac{v^2}{c^2}}$$

Thus the length of the pole relative to S is less than what is relative to S' . But if the pole is rotated in S' and put on the y -axis, then the third equation in (1) implies that $\Delta y = \Delta y'$. Therefore we can compare the lengths of rigid poles in S and S' by putting them across the direction of the motion, but one cannot choose the same scales on all the axes of both S and S' . This implies that the assumption that it is possible to unify scales on all the axes is contradictory.

(5) Suppose that in S' a particle moves along the x' -axis with the velocity

$$\frac{dx'}{dt'} = u'$$

The frame S' moves with velocity v relative to the frame S . Let us find the velocity $u = dx/dt$ with which the particle moves relative to S . To this end, we write equations which are inverse to (1):

$$t = \frac{t' + \frac{v}{c^2} x'}{\sqrt{1 - \frac{v^2}{c^2}}}, \quad x = \frac{vt' + x'}{\sqrt{1 - \frac{v^2}{c^2}}}$$

whence

$$\frac{dx}{dt} = \frac{v + \frac{dx'}{dt'}}{1 + \frac{v}{c^2} \frac{dx'}{dt'}}$$

that is,

$$u = \frac{v + u'}{1 + \frac{v}{c^2} u'} \quad (2)$$

This formula replaces the classical law of addition of velocities, according to which one should have $u = v + u'$ (see § 204 in this connection). Notice that

$$\frac{v+c}{1+\frac{v}{c^2}c} = c$$

which means that by the law of addition of velocities (2) the velocity of light when added to v again yields the velocity of light. This shows that only formula (2), and not the classical law of addition of velocities, is in agreement with the postulate regarding the independence of the velocity of light from the choice of an inertial reference frame.*

* For details concerning mathematical foundations of the theory of relativity see A. S. Eddington: *The Mathematical Theory of Relativity*, Cambridge Univ. Press, Cambridge, 1922.

Part 3

Constant Curvature
Geometry

Chapter 8

Differential Properties of Non-Euclidean Metric

8.1. Metric form of the Euclidean plane

215. The source of all the following derivations is a careful analysis of the structure of the formulas expressing the result of measuring geometrical quantities.

First we will examine the most simple case when the measurement is performed on the Euclidean plane.

Let an orthogonal Cartesian coordinate system with axes Ox and Oy be given on the plane. Then it is known from analytic geometry that the square of the distance Δs between two points $M(x, y)$ and $M_1(x + \Delta x, y + \Delta y)$ is given by the formula

$$\Delta s^2 = \Delta x^2 + \Delta y^2 \quad (1)$$

Furthermore, if $M_2(x + \delta x, y + \delta y)$ is another arbitrary point, the cosine of the angle φ between the segments MM_1 and MM_2 is

$$\cos \varphi = \frac{\Delta x \delta x + \Delta y \delta y}{\sqrt{\Delta x^2 + \Delta y^2} \sqrt{\delta x^2 + \delta y^2}} \quad (2)$$

Finally, the area σ of the rectangle with vertices $P_1(x, y)$, $P_2(x + \Delta x, y)$, $P_3(x, y + \Delta y)$, $P_4(x + \Delta x, y + \Delta y)$ is

$$\sigma = \Delta x \Delta y \quad (3)$$

By means of suitable limiting processes these formulas yield more general formulas of differential geometry, which are applicable to curvilinear images. Namely, if $x = \varphi(t)$, $y = \psi(t)$ are parametric equations of a smooth curve, the length s of its arc, as the parameter t runs from a to b , is given by the integral

$$s = \int_a^b \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} dt$$

and the square of the differential of an arc is

$$ds^2 = dx^2 + dy^2 \quad (I)$$

If dy/dx and $\delta y/\delta x$ are the slopes of two curves at the point of their intersection, then the angle φ between the curves can be found from

$$\cos \varphi = \frac{dx \delta x + dy \delta y}{\sqrt{dx^2 + dy^2} \sqrt{\delta x^2 + \delta y^2}} \quad (II)$$

Finally, if Σ is some squarable region, its area σ is given by

$$\sigma = \int_{\Sigma} dx dy \quad (\text{III})$$

Formulas (I), (II), (III) characterize analytically the law of measuring lengths, angles and areas expressed by the axioms of Euclidean plane geometry. Therefore the above formulas are said to define the Euclidean metric.

In writing formulas (I), (II), (III) we made use of the orthogonal Cartesian coordinate system. Had we used any other coordinate system, the expressions for ds^2 , $\cos \varphi$ and σ would have had another form.

For example, if r, θ are the polar coordinates of any point and $r = \varphi(t)$, $\theta = \psi(t)$ are the equations of some curve, the differential of the arc of this curve corresponding to the given dt can be expressed in terms of the differentials $dr = \varphi'(t) dt$, $d\theta = \psi'(t) dt$ by means of the expression

$$ds^2 = dr^2 + r^2 d\theta^2 \quad (\text{I}')$$

If $dr, d\theta$ are the variations in the polar coordinates of the point (r, θ) when this point undergoes an infinitesimal change along some curve and $\delta r, \delta \theta$ are the variations in the polar coordinates of this point along some other curve, then the cosine of the angle φ between the curves is

$$\cos \varphi = \frac{dr \delta r + r^2 d\theta \delta \theta}{\sqrt{dr^2 + r^2 d\theta^2} \sqrt{\delta r^2 + r^2 \delta \theta^2}} \quad (\text{II}')$$

Finally, if Σ is a squarable region and σ its area,

$$\sigma = \int_{\Sigma} r dr d\theta \quad (\text{III}')$$

Formulas (I'), (II'), (III') define analytically the metric of the Euclidean plane in polar coordinates.

In order to remove the peculiarities which appear in formulas for ds^2 , $\cos \varphi$ and σ on account of the given coordinate system and in order to discover the general structure of these formulas, we shall obtain these formulas in an arbitrary coordinate system. The desired formulas will be obtained on the basis of formulas (I), (II), (III).

Let there be given a coordinate system (u, v) defined by the equations $x = x(u, v)$, $y = y(u, v)$ which enable us to obtain the Cartesian coordinates x, y of a point in the plane if the coordinates u, v of this point are given (on the class of admissible coordinate systems we impose the condition that the functions $x(u, v)$, $y(u, v)$ are continuously differentiable and their Jacobian $\frac{D(x, y)}{D(u, v)} \neq 0$; the last condition guarantees that the relations $x = x(u, v)$, $y =$

$= y(u, v)$ are reversibility in the neighbourhood of any point and that the functions $u = u(x, y)$, $v = v(x, y)$ are continuously differentiable).

Consider a curve $u = u(t)$, $v = v(t)$. If dt is the change in parameter t and ds the differential of the arc of this curve corresponding to dt , then the expression for ds can be obtained in terms of $du = u'(t)dt$, $dv = v'(t)dt$ by substituting the quantities

$$dx = \frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv$$

$$dy = \frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv$$

in the right side of the relation

$$ds^2 = dx^2 + dy^2 \quad (1)$$

Performing this substitution, combining the terms containing du^2 , $du dv$, dv^2 and introducing the notation

$$\left. \begin{aligned} \left(\frac{\partial x}{\partial u} \right)^2 + \left(\frac{\partial y}{\partial u} \right)^2 &= E \\ \frac{\partial x}{\partial u} \frac{\partial x}{\partial v} + \frac{\partial y}{\partial u} \frac{\partial y}{\partial v} &= F \\ \left(\frac{\partial x}{\partial v} \right)^2 + \left(\frac{\partial y}{\partial v} \right)^2 &= G \end{aligned} \right\} \quad (\alpha)$$

we obtain

$$ds^2 = E du^2 + 2F du dv + G dv^2 \quad (I^*)$$

Here the quantities E , F , G , which are determined in the coordinate system (u, v) chosen by equations (α) at every point of the plane, are completely independent of the curve chosen through this point. On the other hand, the differentials du , dv depend exclusively on the curve along which the point having coordinates u , v is displaced. Thus the expression on the right-hand side of (I^*) is a quadratic form in the arguments du , dv with coefficients E , F , G .

Furthermore, if du , dv and δu , δv are differentials of the coordinates u , v corresponding to the displacement of the point in two different directions making an angle φ with each other, by substituting the quantities

$$dx = \frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv, \quad dy = \frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv$$

$$\delta x = \frac{\partial x}{\partial u} \delta u + \frac{\partial x}{\partial v} \delta v, \quad \delta y = \frac{\partial y}{\partial u} \delta u + \frac{\partial y}{\partial v} \delta v$$

in the formula (II)

$$\cos \varphi = \frac{dx \delta x + dy \delta y}{\sqrt{dx^2 + dy^2} \sqrt{\delta x^2 + \delta y^2}}$$

and taking note of (α) , we obtain

$$\cos \varphi = \frac{E du \delta u + F (du \delta v + dv \delta u) + G dv \delta v}{\sqrt{E du^2 + 2F du dv + G dv^2} \sqrt{E \delta u^2 + 2F \delta u \delta v + G \delta v^2}} \quad (\text{II}^*)$$

Finally, a change of variables in integral (III) yields the following expression for area σ of the region Σ :

$$\begin{aligned} \sigma &= \iint_{\Sigma} dx dy = \iint_{\Sigma} \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv = \iint_{\Sigma} \sqrt{\begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}^2} du dv \\ &= \iint_{\Sigma} \sqrt{EG - F^2} du dv \end{aligned} \quad (\text{III}^*)$$

Thus we now have the following three formulas

$$ds^2 = E du^2 + 2F du dv + G dv^2 \quad (\text{I}^*)$$

$$\cos \varphi = \frac{E du \delta u + F (du \delta v + dv \delta u) + G dv \delta v}{\sqrt{E du^2 + 2F du dv + G dv^2} \sqrt{E \delta u^2 + 2F \delta u \delta v + G \delta v^2}} \quad (\text{II}^*)$$

$$\sigma = \iint_{\Sigma} \sqrt{EG - F^2} du dv \quad (\text{III}^*)$$

which represent in any coordinate system the lengths, angles and areas on the Euclidean plane. These formulas contain as particular cases formulas (I)-(III) and (I')-(III').

It is now easy to see that the expressions for $\cos \varphi$ and σ are in a definite manner constructed from the quadratic form ds^2 .

Namely, the numerator in the right side of formula (II*) is a bilinear form obtained by modifying formula (I*) and the expression under the square root in formula (III*) is nothing but the discriminant of form (I*).

Hence the metric of the Euclidean plane is defined in every coordinate system by quadratic form (I*), which on this account is called *metric*.

The researches carried out by mathematicians and mechanicians of the nineteenth century showed that the geometric systems where the measurement of quantities is expressed, as in the Euclidean plane, analytically by differential quadratic form are an extremely common feature of geometry. They are the topics which precisely make up differential geometry. For example, as is known from the theory of surfaces (and which will be recalled in the appropriate place), the lengths, angles and areas of every surface in the Euclidean space are measured according to formulas having exactly the same structure as formulas (I*)-(III*).

Our immediate task will be to prove that the class of geometric systems just mentioned contains non-Euclidean Lobachevskian geometry, that is, in this geometry also the metric is defined by a differential quadratic form.

8.2. The distance between two points on the Lobachevskian plane

§ 216. Since we wish to study the differential-geometric character of the metric for the Lobachevskian plane, we must first of all derive a formula giving the distance between two points in terms of their coordinates (in a convenient coordinate system).

At the root of formula (1) of § 215, which expresses the Euclidean distance between two points in terms of their Cartesian coordinates, lies Pythagoras' theorem.

Pythagoras' theorem is not valid in non-Euclidean Lobachevskian geometry, and to obtain an analogous theorem is a very difficult task; what is more, the coordinate systems which can be introduced on the Lobachevskian plane in analogy with the Cartesian system turn out to be inconvenient in many respects.

Therefore we abandon the path of development of the synthetic theory of Lobachevsky and choose another path which will lead us directly to the solution of the problem posed. We point out at the outset that the Lobachevskian plane will not be examined in isolation but as an object of the Lobachevskian space.

Let α be the plane in question (in the Lobachevskian space) and take a point O on this plane. There are two horospheres touching α at O (they lie on the opposite sides of α); choose any one of them and call it Σ . All of the following discussion is based on the fact that elementary geometry on Σ is Euclidean (see § 47).

We shall now establish some special mapping of α into Σ . Namely, with every point M on α we associate a point M' on Σ where the horosphere is intersected by the line through M which is parallel to the perpendicular to α at O (Fig. 159; the direction of parallelism is on the same side of α as Σ). It is clear that the images of all points of α do not fill up the horosphere completely; let us determine this region.

Consider a line OA in α which passes through O ; let ON be the perpendicular to α at O (the point N lies on the same side of α as Σ). On the bisector of $\angle AON$ mark off a segment l such that to it there corresponds the angle of parallelism $\Pi(l) = \pi/4$; denote by L the endpoint of this segment.

Due to the manner in which l is chosen, the line u , which is perpendicular to the segment OL at the endpoint L and which lies in the plane of the angle AON , is parallel to the ray OA in one direction and to the ray ON in the other (Fig. 159). Since the line u is parallel

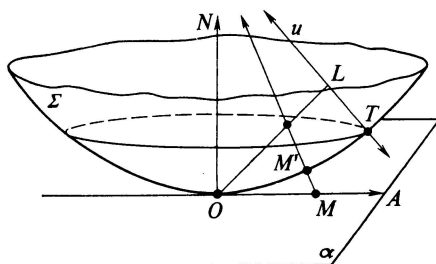


Figure 159

to ON , it is the axis of Σ and hence cuts Σ at some point T . On the ray OA now take any point M (distinct from O) and draw through it a ray parallel to ON . In the plane AON this ray passes between the parallel lines TL and ON . Therefore the point M' at which this ray cuts the oricycle OT lies between the points O and T . On the other hand, if between the points O and T a point P' is taken on the oricycle OT and a line through this point is drawn from O to N parallel to the line ON , then it will be parallel to u in the same direction; in the other direction this line will lean towards the ray OA away from the line u and, accordingly, will cut OA at some point P . This means that every point of the oricycle OT lying between O and T is the image of a certain point of the ray OA . Finally, it is clear that the point O corresponds to itself. Thus the images of all the points of the ray OA fill up the arc of the oricycle OT with one endpoint T excluded.

From this we immediately conclude that the images of all the points of the plane α fill up a region on the horosphere Σ which is "shadowed" by the arc OT (with endpoint T excluded) when it revolves about the line ON .

From the point of view of geometry on a horosphere, this region is nothing but the interior of a disc k with centre O and the radius of which is the arc OT of the oricycle. The arc OT is known as the radius of curvature of the Lobachevskian space. If a certain arc of the oricycle is taken as the unit for measuring lengths on the horosphere Σ , the length of arc OT will be expressed in terms of a number R . The number R (with the scale chosen) will also be called the radius of curvature.

Let us associate with the chosen scale a definite part of the space; for example, let it be the interior of a sphere E whose radius is the distance between the endpoints of the arc of the oricycle which has been chosen as the scale. The quantity $1/R$ can be treated as the "measure of non-Euclideanness" of the space inside E . Namely, the larger the number R the smaller the difference between the Lobachevskian and Euclidean spaces inside E . The precise meaning of

this statement is this: if x is any segment inside E , then the angle of parallelism $\Pi(x)$ tends to $\pi/2$ uniformly as $R \rightarrow \infty$ (see § 230).

In the limiting case $R = \infty$ the disc k is enlarged into the entire horosphere Σ , and then Σ coincides with α and the space becomes Euclidean.

On the horosphere Σ we introduce a rectangular Cartesian coordinate system (x', y') with the origin at O and the scale already chosen. In these coordinates the boundary of k is given by

$$x'^2 + y'^2 = R^2$$

We shall now introduce a coordinate system on the Lobachevskian plane α also. Namely, to every point M of α we assign as coordinates the two numbers

$$x = \frac{x'}{R}, \quad y = \frac{y'}{R}$$

where x', y' are the Cartesian coordinates of the image point M' of the point M on Σ .

It follows from the preceding result that the coordinates of any point of α satisfy the inequality

$$x^2 + y^2 < 1$$

Conversely, if two preassigned numbers x, y satisfy the last inequality, then there is (just one) point on α with the given x, y as its coordinates. The numbers x, y are called the *Beltrami coordinates* of M .

§ 217. We now formulate our problem precisely. Namely we want to derive a formula giving the distance between two points $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ of α in terms of their Beltrami coordinates x_1, y_1 and x_2, y_2 .

For the convenience of the reader the desired formula is derived in stages.

(1) In the preceding subsection we set up a special mapping of the plane α into the interior of the disc k of the horosphere Σ . This mapping possesses the following property, which is the basis of all the conclusions that follow: the images of points of any line in the plane α constitute an arc of oricycle inside the disc k . Indeed, to obtain the image of a point M on α we draw through M a ray parallel to the ray ON up to the point of intersection with the horosphere Σ (see Fig. 159); the point of intersection M' is the desired image of M . Let an arbitrary line a be given in the plane α . Through every point of this line draw a ray parallel to the ray ON ; all the rays drawn lie in one plane λ , parallel to the ray ON , and fill up a strip in this plane. What is more, all the rays drawn are normal to the horosphere Σ ; accordingly, the plane λ cuts the horosphere Σ normally. But the normal section of Σ by λ is an oricycle the part of which

lying in the interior of k constitutes the set of the images of all points of a . Thus the images of all points of an arbitrary line a lying in α constitute an arc of the oricycle, as asserted.

(2) Consider a motion of plane α about itself, that is, a mapping of α into itself such that the distance between any two points of this plane is equal to the distance between their images. This mapping will be expressed by $M^* = \varphi(M)$, where M is any point of α and M^* its image; the mapping of α into the interior of the disc k of horosphere Σ , examined above, will be written $M' = \psi(M)$. The two mappings $M^* = \varphi(M)$ and $M' = \psi(M)$ induce a mapping $M'^* = \chi(M')$ of the interior of k into itself; here M' is any point in the interior of k , point M'^* its image, and $M'^* = \chi(M') = \psi(M^*)$, $M^* = \varphi(M)$, $M = \psi^{-1}(M')$, where ψ^{-1} denotes the inverse of ψ . In other words, under a motion of the points of the plane α their images are displaced on the horosphere Σ ; this very displacement has been expressed by $M'^* = \chi(M')$.

Note now that under the motion $M^* = \varphi(M)$ the points lying on a certain line in α are mapped into points also lying on one line; in short, under the motion of plane α about itself all its lines are mapped into lines. Further, as shown in item 1, under the mapping $M' = \psi(M)$ the points lying on one line in α are mapped into points constituting an arc of an oricycle in the interior of k ; in short, under the mapping $M' = \psi(M)$ of plane α into the disc k the lines of this plane are mapped into arcs of oricycles. Combining these two results, we can conclude that under the mapping $M'^* = \chi(M')$ of the disc k into itself all the arcs of oricycles lying inside k are mapped into arcs of oricycles.

In the sense of elementary geometry of the horosphere Σ , which is a Euclidean geometry, the oricycles are straight lines. Taking note of this fact, the preceding conclusion can be stated thus: under the mapping $M'^* = \chi(M')$ the interior of the disc k is mapped into itself in such a manner that the chords of k again become its chords.

(3) We shall define the cross ratio of four points of an oricycle in exactly the same way as is done for four points of a Euclidean line (see § 137, formula (*)).

Let M'_1, M'_2 be any two points of the horosphere Σ lying in the interior of the disc k and M'^*_1, M'^*_2 their images under the mapping χ . Let P', Q' and P'^*, Q'^* be the points where the oricycles $M'_1 M'_2$ and $M'^*_1 M'^*_2$ cut the disc k and are taken in such a way that the order of points P', Q', M'_1, M'_2 on the oricycle $M'_1 M'_2$ is the same as the order of points $P'^*, Q'^*, M'^*_1, M'^*_2$ on the oricycle $M'^*_1 M'^*_2$. Then

$$(P'^* Q'^* M'^*_1 M'^*_2) = (P' Q' M'_1 M'_2)$$

that is, the cross ratio of the points $P'^*, Q'^*, M'^*_1, M'^*_2$ equals the cross ratio of the points P', Q', M'_1, M'_2 (see § 138).

(4) The last result holds the key to the solution of our problem.

We are looking for a formula which will enable us to calculate the distance between any two points M_1 and M_2 of the plane α if their Beltrami coordinates are known.

Consider the images M'_1 and M'_2 of the points M_1 and M_2 under the mapping $M' = \psi(M)$ of the plane α into the interior of the disc k on the horosphere Σ (see item 2), and through M'_1, M'_2 draw an oricycle on Σ ; let P', Q' be the points of intersection of this oricycle with the boundary of the disc k . Let $(P'Q'M'_1M'_2)$ be the cross ratio of the points P', Q', M'_1, M'_2 defined on Σ in the usual Euclidean sense. We assert that

The distance between any two points M_1 and M_2 of the Lobachevskian plane α is given by the formula

$$\rho(M_1, M_2) = c |\ln (P'Q'M'_1M'_2)| \quad (*)$$

where c is a positive constant which depends on the choice of the scale of measurement.

Proof. First of all note that $\rho(M_1, M_2) > 0$ if M_1 and M_2 are distinct points. Indeed, if M_1 and M_2 are distinct, then $P'M'_1/M'_1Q' \neq P'M'_2/M'_2Q'$; accordingly, $(P'Q'M'_1M'_2) \neq 1$ and $\ln (P'Q'M'_1M'_2) \neq 0$. Next, we shall establish the following facts.

(a) Suppose that under a motion $M^* = \varphi(M)$ of the plane α about itself, the points M_1, M_2 are mapped into the points M_1^*, M_2^* . To the points M_1, M_2, M_1^*, M_2^* of α there correspond the points $M'_1, M'_2, M_1'^*, M_2'^*$ on the horosphere Σ and to the motion $M^* = \varphi(M)$ there corresponds a mapping $M'^* = \chi(M')$ of the disc k into itself under which the points M'_1, M'_2 are mapped into the points $M_1'^*, M_2'^*$. As above, let P', Q' be the points of intersection of the oricycle $M'_1M'_2$ with the boundary of the disc k ; similarly, by means of the points of the oricycle $M_1'^*M_2'^*$ we determine the points P'^*, Q'^* . If the symbols P', Q' and P'^*, Q'^* are taken in a suitably consistent way, then, by item 3 above, we have

$$(P'^*Q'^*M_1'^*M_2'^*) = (P'Q'M'_1M'_2)$$

immediately implying that

$$\rho(M_1^*, M_2^*) = \rho(M_1, M_2)$$

(b) On the line M_1M_2 take a third point M_3 such that M_2 lies between M_1 and M_3 . To the points M_1, M_2, M_3 there correspond the points M'_1, M'_2, M'_3 on the horosphere Σ lying on one oricycle in which M'_2 lies between M'_1 and M'_3 . Let P', Q' have the old meaning, but we assume that on the oricycle $M'_1M'_2M'_3$ the direction of $P'Q'$ is opposite that of $M'_1M'_2M'_3$. Under this condition, we have $P'M'_1/M'_1Q' > P'M'_2/M'_2Q'$; accordingly, $(P'Q'M'_1M'_2) > 1$ and in exactly the same way

$$(P'Q'M'_2M'_3) < 1, \quad (P'Q'M'_1M'_3) < 1$$

We write the relations

$$(P'Q'M'_1M'_3) = \frac{P'M'_1}{M'_1Q'} \div \frac{P'M'_3}{M'_3Q'} = \left(\frac{P'M'_1}{M'_1Q'} \div \frac{P'M'_2}{M'_2Q'} \right) \left(\frac{P'M'_2}{M'_2Q'} \div \frac{P'M'_3}{M'_3Q'} \right) \\ = (P'Q'M'_1M'_2) (P'Q'M'_2M'_3)$$

whence

$$\ln (P'Q'M'_1M'_3) = \ln (P'Q'M'_1M'_2) + \ln (P'Q'M'_2M'_3)$$

Since all the cross ratios present here are greater than 1, it follows that their logarithms are positive and consequently coincide with their absolute values. Thus we can write

$$|\ln (P'Q'M'_1M'_3)| = |\ln (P'Q'M'_1M'_2)| + |\ln (P'Q'M'_2M'_3)|$$

leading to

$$\rho (M_1, M_3) = \rho (M_1, M_2) + \rho (M_2, M_3)$$

(c) Let a segment E_1E_2 be taken as the unit of length. Since E_1, E_2 are distinct points, it follows that $c_1 = |\ln (P'_cQ'_cE'_1E'_2)| > 0$ (here P'_c, Q'_c are the points of intersection of the oricycle $E'_1E'_2$ with the boundary of the disc k). If we set $c = 1/c_1$ in (*), then

$$\rho (E_1, E_2) = 1.$$

Thus according to formula (*) a definite positive number corresponds to every segment, and

(a) to equal segments there correspond equal numbers;

(b) if M_2 is a point of the segment M_1M_3 and the numbers $\rho (M_1, M_2) = a$, $\rho (M_2, M_3) = b$ correspond to the segments M_1M_2 and M_2M_3 , then the number $\rho (M_1, M_3) = a + b$ corresponds to the segment M_1M_3 ;

(c) to a certain segment E_1E_2 there corresponds the number 1.

But we know that these conditions uniquely define the length of a segment (see § 20), hence we have shown that formula (*) defines the length of the segment M_1M_2 .

The main portion of the derivation of the desired formula has been completed; the rest is connected with elementary computations.

(5) As in § 216, we shall denote the Beltrami coordinates of a point M on plane α by x, y and the Cartesian coordinates of its image M' on Σ by x', y' ; we know that then $x' = Rx, y' = Ry$.

Besides the given points $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ on α , the distance between which we have to obtain, consider their images $M'_1(x'_1, y'_1)$ and $M'_2(x'_2, y'_2)$ on Σ . Let $M'(x', y')$ be any point of the oricycle $M'_1M'_2$. Setting $M'_1M'/M'M'_2 = \lambda$ we obtain the known relations

$$x' = \frac{x'_1 + \lambda x'_2}{1 + \lambda}, \quad y' = \frac{y'_1 + \lambda y'_2}{1 + \lambda} \quad (1)$$

Let $\lambda = \lambda_1$ and $\lambda = \lambda_2$ be the values of the parameter λ corresponding to the case when M' occupies the positions of P' and Q' , respectively.

on the boundary of k . Then we have

$$(P'Q'M_1^*M_2') = \frac{P'M_1'}{M_1'Q'} \div \frac{P'M_2'}{M_2'Q'} = \frac{M_1'P'}{P'M_2'} \div \frac{M_2'Q'}{Q'M_1'} = \frac{\lambda_1}{\lambda_2}$$

In order to find λ_1, λ_2 we must substitute the right-hand sides of (1) in the equation of the boundary of k

$$x'^2 + y'^2 = R^2$$

and solve the resulting quadratic equation in λ . Performing this substitution and changing to Beltrami coordinates of the given points, we have

$$(x_1 + \lambda x_2)^2 + (y_1 + \lambda y_2)^2 - (1 + \lambda)^2 = 0$$

Writing

$$x_1^2 + y_1^2 - 1 = \Omega_{11}$$

$$x_1x_2 + y_1y_2 - 1 = \Omega_{12}$$

$$x_2^2 + y_2^2 - 1 = \Omega_{22}$$

for brevity, the last equation can be written in the form

$$\Omega_{22}\lambda^2 + 2\Omega_{12}\lambda + \Omega_{11} = 0$$

whence

$$\lambda_{1,2} = \frac{-\Omega_{12} \pm \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}}{\Omega_{22}}$$

Choosing λ_1, λ_2 appropriately, we have*

$$(P'Q'M_1^*M_2') = \frac{\lambda_1}{\lambda_2} = \frac{\Omega_{12} - \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}}{\Omega_{12} + \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}} < 1$$

Consequently,

$$\rho(M_1, M_2) = c \ln \frac{\Omega_{12} - \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}}{\Omega_{12} + \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}} \quad (2)$$

This is the desired formula. We shall take one more step, that of imposing a definite condition for the choice of the constant c . The thing is that at the proper place a certain scale was assumed to have been established on Σ ; moreover, a scale was defined on α . As long as these two scales are not subject to any condition, the constant c remains undefined.

The choice of scales will be governed by the following condition. Let M be any point of α , point M' its image on Σ ; assume that M approaches O along a line, then M' will approach O along an oricycle. We require that

$$\lim_{M \rightarrow O} \frac{OM}{OM'} = 1$$

* It should be noted here that the quantities $\Omega_{11}, \Omega_{12}, \Omega_{22}$ are negative.

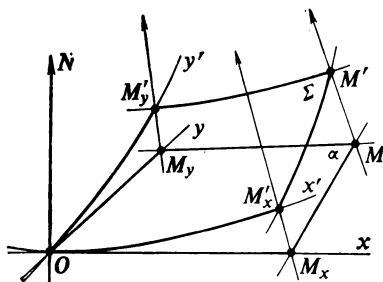


Figure 160

where OM is the length of the line segment and OM' that of the arc of the oricycle. We shall determine c by this condition.

For convenience in computation, assume that M' lies on the Ox' -axis of the Cartesian system introduced on Σ . Then

$$OM' = x' = Rx, \quad OM = \rho(O, M) = c \ln \frac{1+x}{1-x}$$

(we have applied formula (2) by taking $x_1 = 0$, $y_1 = 0$, $x_2 = x$, $y_2 = 0$), and therefore

$$\lim_{M \rightarrow O} \frac{OM}{OM'} = \lim_{x \rightarrow 0} \frac{c}{Rx} \ln \frac{1+x}{1-x} = \frac{2c}{R} = 1$$

hence $c = R/2$.

The final form of the formula giving non-Euclidean distance between two points in terms of their Beltrami coordinates is

$$\rho(M_1, M_2) = \frac{R}{2} \ln \frac{\Omega_{12} - \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}}{\Omega_{12} + \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}} \quad (3)$$

§ 218. We shall now describe the Beltrami coordinate system without using the embracing space.

For the time being we shall use the Cartesian system on the horosphere Σ , as was done in § 216 when introducing the Beltrami coordinates.

Let Ox' and Oy' be two oricycles which serve as the axes of the Cartesian system on the horosphere Σ (Fig. 160). Under the mapping of the plane α onto the horosphere Σ the oricycles Ox' and Oy' have as their preimages two mutually perpendicular lines in plane α passing through O ; we shall call them Ox and Oy . Inside the angle formed by the positive directions of the lines Ox and Oy on α take any point M with Beltrami coordinates x, y ; its image M' on Σ has (positive) Cartesian coordinates $x' = Rx$, $y' = Ry$. From M draw a perpendicular to the line Ox and call the foot of the perpendicular M_x . From M_x draw a ray parallel to the ray ON (ON has

the same significance as in § 216), and let M'_x be its point of intersection with Σ . Now notice that (a) M'_x lies on the oricycle Ox' on its positive side; (b) the points M, M', M_x, M'_x lie in one plane; (c) the line MM_x is perpendicular to the plane NOM_x , and, consequently, the plane $MM'M_xM'_x$ passing through this line is also perpendicular to the plane NOM_x ; (d) the last result enables us to conclude that the arc of the oricycle $M'M'_x$ is perpendicular to the oricycle Ox' . Thus the point M'_x on Σ corresponding to the point M_x is the foot of the perpendicular, in the sense of the geometry of a horosphere, which is drawn from M' to the axis Ox' . Similarly one obtains the points M_y and M'_y on Oy and Oy' (the construction is shown in Fig. 160). Thus we have

$$OM'_x = x' = Rx, \quad OM'_y = y' = Ry$$

Setting

$$OM_x = \xi, \quad OM_y = \zeta$$

we shall express these quantities in terms of x and y . We shall find ξ , and ζ can be found similarly. Since we know the Cartesian coordinates of M'_x —they are $x', 0$ —we also know the Beltrami coordinates of M_x —they are $x, 0$. Therefore to obtain ξ it is enough to use formula (3), § 217, with $x_1 = 0, y_1 = 0, x_2 = x, y_2 = 0$. This yields

$$\xi = \frac{R}{2} \ln \frac{1+x}{1-x}$$

Similarly,

$$\zeta = \frac{R}{2} \ln \frac{1+y}{1-y}$$

These relations yield the desired expression for the Beltrami coordinates:

$$x = \tanh \frac{\xi}{R}, \quad y = \tanh \frac{\zeta}{R} \quad (1)$$

In deriving these formulas x, y, ξ, ζ were assumed to be positive, but if ξ and ζ denote the signed segments OM_x and OM_y in accordance with the usual rule, then formulas (1) express the Beltrami coordinates for any position of M . This is because the hyperbolic tangent is an odd function: $\tanh(-\alpha) = -\tanh \alpha$.

We are now in a position to describe the Beltrami coordinates without using the space embracing the plane; two mutually perpendicular axes Ox and Oy and a scale are chosen on the plane; from any point M perpendiculars MM_x and MM_y are drawn to the axes Ox and Oy , respectively (Fig. 161). Thus two numbers $OM_x = \xi$ and $OM_y = \zeta$ are determined. The Beltrami coordinates of M are the numbers

$$x = \tanh \frac{\xi}{R}, \quad y = \tanh \frac{\zeta}{R}$$

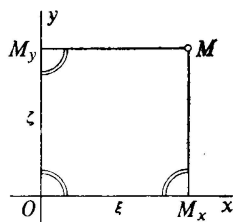


Figure 161

If on the basis of the preassigned Beltrami coordinates x, y we wish to obtain the corresponding point, we must find the segments ξ and ζ first and then mark off these segments on the respective axes and, finally, erect perpendiculars to the axes at the endpoints of these segments; the point with given coordinates x, y is the intersection of these perpendiculars.

The perpendiculars just mentioned will intersect if and only if the given numbers x, y satisfy the inequality

$$x^2 + y^2 < 1$$

we know that this condition is both necessary and sufficient in order for the numbers x, y to be the Beltrami coordinates for any point of the Lobachevskian plane (see § 216).

A striking feature is that a line in Beltrami coordinates is defined by a first-degree equation. Indeed, let u be a line in the plane α and M a variable point on this line with the current Beltrami coordinates x, y . On the horosphere Σ the image of u is the oricycle u' and the image of M is the point M' with the Cartesian coordinates x', y' . Because in geometry on a horosphere the oricycle u' plays the role of a line, this can be represented in Cartesian coordinates by a first-degree equation

$$A'x' + B'y' + C' = 0$$

Setting in this equation $x' = Rx, y' = Ry$ and, further, $A'R = A, B'R = B, C' = C$, we obtain the equation of the given line

$$Ax + By + C = 0$$

which is a first-degree equation (subject to the additional condition $x^2 + y^2 < 1$).

§ 219. Somewhat later we shall have to use a transformation of Beltrami coordinates.

Suppose that two systems of Beltrami coordinates (with the same scale, for simplicity) are given on the Lobachevskian plane. Any point M of the plane has the coordinates (x, y) in one of these systems and $(\bar{x}, \bar{y})$ in the other. The quantities $(\bar{x}, \bar{y})$ are functions of x, y .

It will prove important for us that these functions (1) are continuously differentiable, and (2) have a nonzero Jacobian.

We shall prove these facts. Let us perform a motion of the plane about itself so that the new coordinate axes coincide with the old. Then M is mapped into the point $M^* = \varphi(M)$. It is easy to see that the old coordinates (x^*, y^*) of M^* coincide with the new coordinates of M . Thus

$$x^* = \bar{x}, \quad y^* = \bar{y}$$

Let us examine once more the already known mapping of the plane onto the horosphere Σ which touches the plane at the origin of the old coordinate system. The motion $M^* = \varphi(M)$ induces a mapping of the interior of the disc k on the horosphere Σ into itself; as before, this will be denoted by $M'^* = \chi(M')$.

The mapping $M'^* = \chi(M')$ transforms chords of k into chords of this disc. This together with the result of § 138 enables us to conclude that this mapping in the Cartesian system (x', y') on Σ , which corresponds to the Beltrami system (x, y) on the given plane, is expressed by formulas of the form

$$\left. \begin{aligned} x'^* &= \frac{a'_1 x' + b'_1 y' + c'_1}{\alpha' x' + \beta' y' + \gamma'} \\ y'^* &= \frac{a'_2 x' + b'_2 y' + c'_2}{\alpha' x' + \beta' y' + \gamma'} \end{aligned} \right\} \quad (*)$$

with the condition that

$$\Delta = \begin{vmatrix} a'_1 & b'_1 & c'_1 \\ a'_2 & b'_2 & c'_2 \\ \alpha' & \beta' & \gamma' \end{vmatrix} \neq 0$$

Using the relations $x' = Rx$, $y' = Ry$, $x'^* = Rx^* = R\bar{x}$, $y'^* = Ry^* = R\bar{y}$, formulas (*) imply that

$$\left. \begin{aligned} \bar{x} &= \frac{a'_1 Rx + b'_1 Ry + c'_1}{\alpha' R^2 x + \beta' R^2 y + \gamma' R} \\ \bar{y} &= \frac{a'_2 Rx + b'_2 Ry + c'_2}{\alpha' R^2 x + \beta' R^2 y + \gamma' R} \end{aligned} \right\} \quad (**)$$

If the new notation

$$\begin{aligned} a'_1 R &= a_1, & b'_1 R &= b_1, & c'_1 &= c_1 \\ a'_2 R &= a_2, & b'_2 R &= b_2, & c'_2 &= c_2 \\ \alpha' R^2 &= \alpha, & \beta' R^2 &= \beta, & \gamma' R &= \gamma \end{aligned}$$

for the coefficients are introduced, then formulas (**) assume the form

$$\left. \begin{aligned} \bar{x} &= \frac{a_1x + b_1y + c_1}{\alpha x + \beta y + \gamma} \\ \bar{y} &= \frac{a_2x + b_2y + c_2}{\alpha x + \beta y + \gamma} \end{aligned} \right\} \quad (***)$$

with

$$\Delta = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ \alpha & \beta & \gamma \end{vmatrix} \neq 0$$

(because $\Delta = R^3\Delta'$ and $\Delta' \neq 0$). Formulas (***) express the new Beltrami coordinates of any point M in terms of its old Beltrami coordinates (the coefficients $a_1, b_1, \dots, \gamma$ in these formulas depend on how the new axes are situated with respect to the old axes). We shall now demonstrate that the formulas (***) are meaningful for all admissible values of Beltrami coordinates, that is, for all values of x, y satisfying $x^2 + y^2 < 1$. Indeed, if $\alpha x_0 + \beta y_0 + \gamma = 0$ for $x_0^2 + y_0^2 < 1$, then for those values of x, y which are sufficiently close to x_0, y_0 and satisfy $x^2 + y^2 < 1$, $\alpha x + \beta y + \gamma \neq 0$ the corresponding values of $\bar{x}, \bar{y}$ would be as large as possible, which is impossible because $\bar{x}^2 + \bar{y}^2 < 1$.

Thus $\bar{x}, \bar{y}$ are expressed in terms of x, y by means of rational fractions the denominators of which are not zero for all admissible values of x, y . This implies that $\bar{x}, \bar{y}$ are continuously differentiable with respect to x, y at all points of the Lobachevskian plane. Further, simple computations yield

$$\frac{D(\bar{x}, \bar{y})}{D(x, y)} = \frac{\Delta}{(\alpha x + \beta y + \gamma)^3}$$

showing that at all points of the Lobachevskian plane the functions $\bar{x}, \bar{y}$ have a nonzero Jacobian.

8.3. Metric form of the Lobachevskian plane

§ 220. We are now in a position to establish the principal result that was stated in § 217: show that the metric for the Lobachevskian plane is defined by a differential quadratic form. To this end, we shall obtain the expression giving the length of the arc of any smooth curve. At the foundation of the following computations lies formula (3), § 217, expressing the distance between two points of the Lobachevskian plane in terms of their Beltrami coordinates. This formula will now be put in a special form.

Let M_1, M_2 be two points of the Lobachevskian plane, and let $(x, y), (x + \Delta x, y + \Delta y)$, respectively, be their Beltrami coordinates and $\Delta\rho$ the length of the segment M_1M_2 . The quantities Δx and Δy are assumed to be infinitesimally small. The quantities $\Omega_{11}, \Omega_{12}, \Omega_{22}$ present in the right side of formula (3), § 217, now assume the form

$$\left. \begin{aligned} \Omega_{11} &= x^2 + y^2 - 1 \\ \Omega_{12} &= x^2 + y^2 - 1 + x \Delta x + y \Delta y \\ \Omega_{22} &= x^2 + y^2 - 1 + 2x \Delta x + 2y \Delta y + \Delta x^2 + \Delta y^2 \end{aligned} \right\} \quad (1)$$

All these quantities are negative because of the basic relation $x^2 + y^2 < 1$ connecting the Beltrami coordinates of any point. Relations (1) yield

$$\begin{aligned} \Omega_{12}^2 - \Omega_{11}\Omega_{22} &= (\Omega_{12} - \Omega_{11})^2 - \Omega_{11}(\Omega_{22} - 2\Omega_{12} + \Omega_{11}) \\ &= (x \Delta x + y \Delta y)^2 + (1 - x^2 - y^2)(\Delta x^2 + \Delta y^2) \\ &= (1 - y^2)\Delta x^2 + 2xy \Delta x \Delta y + (1 - x^2)\Delta y^2 \end{aligned} \quad (2)$$

showing that $\Omega_{12}^2 - \Omega_{11}\Omega_{22}$ is a positive quantity which is infinitesimally small together with $\Delta x, \Delta y$.

Applying formula (3), § 217, we obtain

$$\begin{aligned} \Delta\rho &= \frac{R}{2} \ln \frac{\Omega_{12} - \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}}{\Omega_{12} + \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}} = \frac{R}{2} \ln \left(1 - \frac{2\sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}}{\Omega_{12} + \sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}} \right) \\ &= R \frac{\sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}}{1 - x^2 - y^2} + \alpha(x, y, \Delta x, \Delta y) \end{aligned} \quad (3)$$

where α is an infinitesimally small quantity of a higher order in comparison to $\sqrt{\Omega_{12}^2 - \Omega_{11}\Omega_{22}}$. Relations (3) and (2) yield the desired special form for formula (3) of § 217:

$$\Delta\rho = R \frac{\sqrt{(1-y^2)\Delta x^2 + 2xy \Delta x \Delta y + (1-x^2)\Delta y^2}}{1 - x^2 - y^2} + \alpha(x, y, \Delta x, \Delta y) \quad (4)$$

from which the expression for the length of an arc can be easily obtained.

Suppose that a curve is defined by the parametric equations

$$x = x(t), \quad y = y(t)$$

When t changes on the segment $t_0 \leq t \leq T$ the variable point describes an arc A_0A of the curve so that the variable point coincides with the initial point A_0 of the arc if $t_0 = t$ and with the final point A if $t = T$. If the functions $x(t)$ and $y(t)$ are continuous on $t_0 \leq t \leq T$ together with their first order derivatives and if the derivatives do

not vanish simultaneously at any point of this segment, then the arc A_0A will be called smooth.*

It is important to establish that the smoothness of a curve does not depend on the Beltrami coordinate system chosen. To do this, consider any transformation of the given Beltrami coordinates into new Beltrami coordinates. Let (x, y) and $(\bar{x}, \bar{y})$ be old and new coordinates respectively of any point of the plane; $\bar{x}, \bar{y}$ are functions of x, y :

$$\bar{x} = \bar{x}(x, y), \quad \bar{y} = \bar{y}(x, y)$$

As established in § 219, these functions are continuously differentiable and their Jacobian does not vanish. This implies the existence and continuity of the derivatives $\bar{dx}/dt, \bar{dy}/dt$:

$$\frac{\bar{dx}}{dt} = \frac{\partial \bar{x}}{\partial x} \frac{dx}{dt} + \frac{\partial \bar{x}}{\partial y} \frac{dy}{dt}, \quad \frac{\bar{dy}}{dt} = \frac{\partial \bar{y}}{\partial x} \frac{dx}{dt} + \frac{\partial \bar{y}}{\partial y} \frac{dy}{dt}$$

Further if dx/dt and dy/dt do not vanish simultaneously, it follows that $\bar{dx}/dt$ and $\bar{dy}/dt$ also do not vanish simultaneously, because

$$\begin{vmatrix} \frac{\partial \bar{x}}{\partial x} & \frac{\partial \bar{x}}{\partial y} \\ \frac{\partial \bar{y}}{\partial x} & \frac{\partial \bar{y}}{\partial y} \end{vmatrix} \neq 0$$

Therefore the smoothness property which was assumed to be fulfilled in the (x, y) system will also be fulfilled in any other system $(\bar{x}, \bar{y})$.

In defining the length of the arc A_0A we proceed along the same lines as is done in defining the length of the arc of the curve in Euclidean geometry. We divide the segment $t_0 \leq t \leq T$ in any manner by the points $t_1, \dots, t_n$ arranged in increasing order:

$$t_0 < t_1 < t_2 < t_3 < \dots < t_n = T$$

To every point t_i there corresponds a point A_i on the arc A_0A . Construct a polygonal line $A_0A_1A_2 \dots A_{n-1}A$ and denote its length by σ . Thus to every division of the segment $t_0 \leq t \leq T$ there corresponds a positive number σ , the length of the corresponding polygonal line.

Suppose now that any sequence of divisions of the segment $t_0 \leq t \leq T$ is taken such that the maximum length of the parts of the segment tends to zero. If the corresponding sequence of numbers σ tends to a definite limit s which is independent of the sequence of divisions of $t_0 \leq t \leq T$ chosen, then this limit, that is, the number s will be called the length of the arc A_0A .

* The conditions $\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 \neq 0$ here has the same sense as in Euclidean differential geometry.

Setting

$$A_i A_{i+1} = \Delta \rho_i$$

$$x(t_{i+1}) - x(t_i) = \Delta x_i, \quad y(t_{i+1}) - y(t_i) = \Delta y_i$$

and using formula (4), the length of a polygonal line is given by

$$\sigma = \sum_{i=0}^{n-1} \Delta \rho_i = R \sum_{i=0}^{n-1} \frac{\sqrt{(1-y_i^2)(\Delta x_i)^2 + 2x_i y_i \Delta x_i \Delta y_i + (1-x_i^2)(\Delta y_i)^2}}{(1-x_i^2 - y_i^2)^{3/2}} + \sum_{i=0}^{n-1} \alpha(x_i, y_i, \Delta x_i, \Delta y_i)$$

and passing to limit we obtain

$$s = R \int_{t_0}^T \frac{\sqrt{(1-y^2)x'^2 + 2xyx'y' + (1-x^2)y'^2}}{1-x^2-y^2} dt \quad (5)$$

where the dashes denote differentiation with respect to t .*

The integral in the right side of (5) exists because the functions $x(t)$, $y(t)$, $x'(t)$, $y'(t)$ are continuous. If the upper limit in the integral of formula (5) is variable, then this formula expresses the arc of the curve as a function of T :

$$s = s(T)$$

The differential of the arc is given by

$$ds = R \frac{\sqrt{(1-y^2)dx^2 + 2xy dx dy + (1-x^2)dy^2}}{1-x^2-y^2}$$

or

$$ds^2 = R^2 \frac{(1-y^2)dx^2 + 2xy dx dy + (1-x^2)dy^2}{(1-x^2-y^2)^2} \quad (6)$$

Thus the square of the differential of the arc is a quadratic form in the differentials dx and dy . If we set

$$\frac{R^2(1-y^2)}{(1-x^2-y^2)^2} = E(x, y), \quad \frac{R^2xy}{(1-x^2-y^2)^2} = F(x, y)$$

$$\frac{R^2(1-x^2)}{(1-x^2-y^2)^2} = G(x, y) \quad (7)$$

* To prove that $\sum \alpha(x_i, y_i, \Delta x_i, \Delta y_i) \rightarrow 0$, it is enough to notice that $\alpha_i = \alpha(x_i, y_i, \Delta x_i, \Delta y_i)$ is a small quantity of the second order in comparison to $\Delta t_i = t_{i+1} - t_i$. More precisely, if the entire curve lies in the domain $1 - x^2 - y^2 > \varepsilon > 0$, then the ratio $|\alpha_i|/\Delta t_i^2$ is bounded above by a number that depends on ε and on the upper bounds for the quantities $|x'(t)|$, $|y'(t)|$. This follows from the definition of α_i according to formula (3) and from the expressions for Δx_i , Δy_i in terms of Δt_i according to Lagrange's formula.

formula (6) can be written in the form

$$ds^2 = E dx^2 + 2F dx dy + G dy^2 \quad (8)$$

The quadratic form $E dx^2 + 2F dx dy + G dy^2$ defines the length of curves in the Lobachevskian plane. On this account it will be called the *metric form for the Lobachevskian plane*.

§ 221. We shall now derive a formula for the angle between two curves. As shown in § 218, the equation

$$Ax + By + C = 0 \quad (1)$$

represents a line in the Beltrami coordinates. If a point $M_1(x_1, y_1)$ lies on this line, its coordinates must satisfy equation (1), that is, we must have

$$Ax_1 + By_1 + C = 0 \quad (2)$$

Subtracting (2) from (1), we obtain

$$y - y_1 = k(x - x_1)$$

where $k = -A/B$.

The quantity k in the last equation will be called the *direction parameter of the line*.

The angle between any two curves is naturally defined as the angle between their tangents. Since a line in Beltrami coordinates is represented by a first degree equation, it follows that the equation of the tangent in these coordinates has exactly the same form as the equation of the tangent in the Cartesian coordinates in the Euclidean plane, and its direction parameter is expressed in exactly the same way as the slope of the tangent in Euclidean geometry.

Indeed, let a curve be given in the parametric form:

$$x = \varphi(t), \quad y = \psi(t)$$

On this curve take two points M and M' corresponding to the two values t and t' of the parameter. The equation of the secant MM' in Beltrami coordinates is evidently of the form

$$\frac{X - \varphi(t)}{\varphi(t') - \varphi(t)} = \frac{Y - \psi(t)}{\psi(t') - \psi(t)}$$

Dividing the denominators of both sides by $t' - t$ and then taking the limits as $t' \rightarrow t$, the equation of the tangent is obtained in the form

$$\frac{X - \varphi(t)}{\varphi'(t)} = \frac{Y - \psi(t)}{\psi'(t)}$$

and the direction parameter

$$k = \frac{\psi'(t)}{\varphi'(t)} = \frac{dy}{dx}$$

Let us take any two curves whose directions of which at the common point $M(x, y)$ are given by the parameters $k_1 = dy/dx$, $k_2 = \delta y/\delta x$. We shall prove that the angle φ between these curves is expressed by a formula with exactly the same structure as formula (II*) of § 215, giving the Euclidean angle in an arbitrary coordinate system:

$$\cos \varphi = \frac{E dx \delta x + F(dx \delta y + dy \delta x) + G dy \delta y}{\sqrt{E dx^2 + 2F dx dy + G dy^2} \sqrt{E \delta x^2 + 2F \delta x \delta y + G \delta y^2}} \quad (3)$$

We shall first establish that the right side of this formula remains invariant under a transformation of Beltrami coordinates.

Consider a new Beltrami coordinate system $(\bar{x}, \bar{y})$, besides the coordinate system (x, y) , whose origin and direction of axes are arbitrary. We represent the metric formula for the Lobachevskian plane in new coordinates:

$$ds^2 = \bar{E} d\bar{x}^2 + 2\bar{F} d\bar{x} d\bar{y} + \bar{G} d\bar{y}^2$$

Since ds^2 does not depend on the coordinate system, we must have the identity

$$\bar{E} d\bar{x}^2 + 2\bar{F} d\bar{x} d\bar{y} + \bar{G} d\bar{y}^2 = E dx^2 + 2F dx dy + G dy^2 \quad (4)$$

Assuming that the old and new coordinates are connected by the formulas

$$x = x(\bar{x}, \bar{y}), \quad y = y(\bar{x}, \bar{y})$$

and replacing x and y in the right side of identity (4) by their expressions in terms of $\bar{x}$, $\bar{y}$, we obtain

$$\begin{aligned} \bar{E}' d\bar{x}^2 + 2\bar{F}' d\bar{x} d\bar{y} + \bar{G}' d\bar{y}^2 &= \left[E \left(\frac{\partial x}{\partial \bar{x}} \right)^2 + 2F \frac{\partial x}{\partial \bar{x}} \frac{\partial y}{\partial \bar{x}} + G \left(\frac{\partial y}{\partial \bar{x}} \right)^2 \right] d\bar{x}^2 \\ &+ 2 \left[E \frac{\partial x}{\partial \bar{x}} \frac{\partial x}{\partial \bar{y}} + F \left(\frac{\partial x}{\partial \bar{x}} \frac{\partial y}{\partial \bar{y}} + \frac{\partial x}{\partial \bar{y}} \frac{\partial y}{\partial \bar{x}} \right) + G \frac{\partial y}{\partial \bar{x}} \frac{\partial y}{\partial \bar{y}} \right] d\bar{x} d\bar{y} \\ &+ \left[E \left(\frac{\partial x}{\partial \bar{y}} \right)^2 + 2F \frac{\partial x}{\partial \bar{y}} \frac{\partial y}{\partial \bar{y}} + G \left(\frac{\partial y}{\partial \bar{y}} \right)^2 \right] d\bar{y}^2 \end{aligned}$$

But the quantities $d\bar{x}$, $d\bar{y}$, being differentials of independent variables, can assume arbitrary values; therefore the coefficients of the differential quadratic form on the left side of the last identity are equal to the corresponding coefficients on the right side of the same identity:

$$\left. \begin{aligned} \bar{E} &= E \left(\frac{\partial x}{\partial \bar{x}} \right)^2 + 2F \frac{\partial x}{\partial \bar{x}} \frac{\partial y}{\partial \bar{x}} + G \left(\frac{\partial y}{\partial \bar{x}} \right)^2 \\ \bar{F} &= E \frac{\partial x}{\partial \bar{x}} \frac{\partial x}{\partial \bar{y}} + F \left(\frac{\partial x}{\partial \bar{x}} \frac{\partial y}{\partial \bar{y}} + \frac{\partial x}{\partial \bar{y}} \frac{\partial y}{\partial \bar{x}} \right) + G \frac{\partial y}{\partial \bar{x}} \frac{\partial y}{\partial \bar{y}} \\ \bar{G} &= E \left(\frac{\partial x}{\partial \bar{y}} \right)^2 + 2F \frac{\partial x}{\partial \bar{y}} \frac{\partial y}{\partial \bar{y}} + G \left(\frac{\partial y}{\partial \bar{y}} \right)^2 \end{aligned} \right\} \quad (5)$$

The numerator in the right side of formula (3) is a bilinear form

$$E dx \delta x + F (dx \delta y + dy \delta x) + G dy \delta y$$

that is, a homogeneous form linear in the two systems of variables dx, dy and $\delta x, \delta y$. It is easy to see that this form remains invariant under a transformation of Beltrami coordinates, that is

$$\begin{aligned} \bar{E} \bar{dx} \bar{\delta x} + \bar{F} (\bar{dx} \bar{\delta y} + \bar{dy} \bar{\delta x}) + \bar{G} \bar{dy} \bar{\delta y} \\ = E dx \delta x + F (dx \delta y + dy \delta x) + G dy \delta y \end{aligned} \quad (5a)$$

Indeed, replacing the quantities $\bar{E}, \bar{F}, \bar{G}$ on the left side by their expressions given by (5), using the relations

$$dx = \frac{\partial x}{\partial \bar{x}} \bar{dx} + \frac{\partial x}{\partial \bar{y}} \bar{dy}, \quad dy = \frac{\partial y}{\partial \bar{x}} \bar{dx} + \frac{\partial y}{\partial \bar{y}} \bar{dy} \quad (5b)$$

$$\delta x = \frac{\partial x}{\partial \bar{x}} \bar{\delta x} + \frac{\partial x}{\partial \bar{y}} \bar{\delta y}, \quad \delta y = \frac{\partial y}{\partial \bar{x}} \bar{\delta x} + \frac{\partial y}{\partial \bar{y}} \bar{\delta y} \quad (5c)$$

and performing some simple computations, we obtain the right side of formula (5a) and show the validity of this formula. However, identity (5a) can be obtained very easily without using the expressions for $\bar{E}, \bar{F}, \bar{G}$ in terms of E, F, G under condition (4); that is, without using formulas (5). The thing is that relation (4) holds identically, as a consequence of formula (5b), for any $\bar{dx}, \bar{dy}$. Therefore (or, if one likes, as a consequence of formula (5c))

$$\bar{E} \bar{\delta x}^2 + 2\bar{F} \bar{dx} \bar{\delta y} + \bar{G} \bar{\delta y}^2 = E \delta x^2 + 2F \delta x \delta y + G \delta y^2 \quad (4a)$$

$$\begin{aligned} \bar{E} (\bar{dx} + \bar{\delta x})^2 + 2\bar{F} (\bar{dx} + \bar{\delta x}) (\bar{dy} + \bar{\delta y}) + \bar{G} (\bar{dy} + \bar{\delta y})^2 \\ = E (dx + \delta x)^2 + 2F (dx + \delta x) (dy + \delta y) + G (dy + \delta y)^2 \end{aligned} \quad (4b)$$

Subtraction of (4) and (4a) from (4b) results in (5a).

We have proved the invariance of the numerator in the right side of formula (3). The invariance of the denominator is expressed by the relations (4) and (4a).

Thus the right side of formula (3) remains invariant under a change of Beltrami coordinates.

Suppose that we are given a point M with Beltrami coordinates x, y and two curves passing through M having direction parameters

$$k_1 = \frac{dy}{dx}, \quad k_2 = \frac{\delta y}{\delta x}$$

In order to show that (3) gives the angle φ between the given curves, we introduce a new system of Beltrami coordinates with M as the origin. In the new coordinate system E, F, G will have new values as will the direction parameters of the given curves (the ratios of differentials of new coordinates) but the quantity in the right side

of (3) remains unchanged. We avoid making matters more complicated by retaining the old symbols for the quantities in question.

We now have $x = 0, y = 0$ at M . Application of formula (7), § 220, yields the following values for E, F, G at M :

$$E = R^2, \quad F = 0, \quad G = R^2$$

and formula (3) assumes the form

$$\cos \varphi = \frac{dx \delta x + dy \delta y}{\sqrt{dx^2 + dy^2} \sqrt{\delta x^2 + \delta y^2}} \quad (*)$$

Consider the horosphere Σ which touches our plane at M , that is, at the origin of the new coordinate system; we map the plane onto the horosphere as we did in § 216. To every point of the plane with Beltrami coordinates (x, y) there corresponds a point on Σ with the Cartesian coordinates (x', y') , and

$$x = \frac{x'}{R}, \quad y = \frac{y'}{R}$$

Replacing the arguments in the right side of formula (*) by the last formulas, we obtain

$$\cos \varphi = \frac{dx' \delta x' + dy' \delta y'}{\sqrt{dx'^2 + dy'^2} \sqrt{\delta x'^2 + \delta y'^2}} \quad (**)$$

This formula coincides with Euclidean formula (II), § 215, making clear that this gives the angle between the images on Σ of two given curves. But at the point of contact of Σ with our plane the angle between any two curves in the plane equals the angle between their images on the horosphere. Consequently, formula (**), and hence formula (*), gives the angle between the two given curves on the Lobachevskian plane. Thus it is proved that at any point and in any Beltrami coordinates the angle is given by formula (3).

§ 222. Finally, we examine the question of determining areas.

In the Lobachevskian plane we shall consider all possible finite regions which are bounded by smooth closed or piecewise smooth closed curves. Assume that with every such region a positive number is associated such that

(a) equal numbers are associated with congruent regions;

(b) if the region Ω is divided by a piecewise smooth curve into two regions Ω_1 and Ω_2 , the number associated with Ω equals the sum of the numbers associated with Ω_1 and Ω_2 .

In other words, one can say that a positive function of region

$$\sigma = f(\Omega)$$

is defined which (a) assumes equal values on congruent regions, (b) has the additive property, that is,

$$f(\Omega_1 + \Omega_2) = f(\Omega_1) + f(\Omega_2)$$

(here $\Omega_1 + \Omega_2$ denotes a region composed of the points of Ω_1 and Ω_2 as well as the points of the dividing curve).

We shall also define the concept of continuity for a function of region, but to do so we must first define the convergence of sequences of regions.

With every point of the bounded region Ω as centres draw discs of radius ε . The set of interior points of all such discs will be called an ε -neighbourhood of Ω . The ε -neighbourhood of the boundary of the region is similarly defined. Let Ω'_ε denote the set of all points of Ω except those which belong to the ε -neighbourhood of its boundary. Suppose that an infinite sequence of bounded regions $\Omega_1, \Omega_2, \dots, \Omega_n, \dots$ is given. We shall say that the sequence $\Omega_1, \Omega_2, \dots, \Omega_n, \dots$ converges to the region Ω if for any $\varepsilon > 0$ a number N can be found such that for any $n \geq N$ the region Ω_n is enclosed in the ε -neighbourhood of Ω and it itself encloses Ω'_ε .

The region function $f(\Omega)$ will naturally be called continuous if for any region Ω and a sequence of regions Ω_n converging to it the relation

$$\lim_{n \rightarrow \infty} f(\Omega_n) = f(\Omega)$$

holds. For the positive region function conditions (a) and (b) imply its continuity, but we shall not prove this statement. In order to simplify things we shall simply assume that the positive region functions examined below are continuous.

The area of the region Ω in the Lobachevskian plane is the value assumed on this region by a positive function $f(\Omega)$ satisfying conditions (a) and (b).

Here it will be natural to raise the question as to how far conditions (a) and (b) determine a positive function $f(\Omega)$. This question is answered by the following theorem.

Theorem. *If $\varphi(\Omega)$ is a positive region function satisfying conditions (a) and (b), then any other positive region function satisfying the same conditions can be expressed in the form $k\varphi(\Omega)$, where k is a positive constant.*

Thus, according to the definition just given, the areas of all the regions are determined to within a constant factor. This factor can be fixed if a certain region Ω_0 is given a unit area; then the area of any region can be put in the form

$$f(\Omega) = \frac{\varphi(\Omega)}{\varphi(\Omega_0)}$$

where $\varphi(\Omega)$ is an arbitrary region function satisfying conditions (a) and (b).

Let us prove the above theorem. Note that the propositions of § 48 imply the validity of this theorem for triangles; namely, if $f(\Omega)$ is a positive region function satisfying conditions (a) and (b), Δ is

a triangle and $D(\Delta)$ the defect of this triangle, then

$$f(\Delta) = k'D(\Delta) \quad (1)$$

where the constant k' does not depend on the choice of Δ . If $\varphi(\Omega)$ is any other positive region function also satisfying conditions (a) and (b), then, similarly,

$$\varphi(\Delta) = k''D(\Delta) \quad (2)$$

If we set

$$k'/k'' = k$$

we have from (1) and (2)

$$f(\Delta) = k\varphi(\Delta) \quad (3)$$

Evidently a similar relation connects the values assumed by the functions $f(\Omega)$ and $\varphi(\Omega)$ on arbitrary polygons. Indeed, taking any polygon S , divide this in some manner into triangles $\Delta_1, \Delta_2, \dots, \Delta_n$:

$$S = \Delta_1 + \Delta_2 + \dots + \Delta_n$$

Applying relation (3) to triangles:

$$f(\Delta_1) = k\varphi(\Delta_1), \dots, f(\Delta_n) = k\varphi(\Delta_n)$$

and summing up the results obtained, we find that

$$f(\Delta_1) + f(\Delta_2) + \dots + f(\Delta_n) = k[\varphi(\Delta_1) + \dots + \varphi(\Delta_n)]$$

But because the functions f and φ are additive, the last relation can be put in the form

$$f(\Delta_1 + \Delta_2 + \dots + \Delta_n) = k\varphi(\Delta_1 + \dots + \Delta_n)$$

or

$$f(S) = k\varphi(S)$$

Suppose that Ω is now an arbitrary region. We choose a sequence of polygons $\Omega_1, \Omega_2, \dots, \Omega_n, \dots$ converging to Ω in the above defined sense (that such a choice is possible will not be proved here). As shown just now, $f(\Omega_n) = k\varphi(\Omega_n)$ for each of these polygons. Passing to limit as $n \rightarrow \infty$ and noting that f and φ are continuous functions, we find that

$$f(\Omega) = k\varphi(\Omega)$$

that is, conditions (a) and (b) indeed determine a positive region function to within a constant factor. The existence of a function with these properties remains to be proved.

We shall demonstrate that the double integral

$$f(\Omega) = \int_{\Omega} \int \sqrt{EG - F^2} \, dx \, dy \quad (4)$$

is a continuous positive region function satisfying conditions (a) and (b).

First of all notice that because of the basic relation $x^2 + y^2 < 1$ for Beltrami coordinates, the integrand

$$\sqrt{EG - F^2} = \frac{R^2}{(1 - x^2 - y^2)^{3/2}}$$

is positive and continuous at all points of the Lobachevskian plane. This implies that the integral on the right side of (4) exists for any bounded region Ω and has a positive value.

Further, if the functions E , F and G are fixed, that is, if a definite Beltrami coordinate system is chosen, then the value of integral (4) is determined only by the region of integration. What is important, however, is that in reality this value does not depend on the Beltrami coordinate system chosen. To prove this, consider, apart from the given Beltrami system (x, y) , a new Beltrami system $(\bar{x}, \bar{y})$, and let E, F, G and $\bar{E}, \bar{F}, \bar{G}$ be the coefficients of the metric form of the Lobachevskian plane in the old and new coordinates, respectively. Using formulas (5), § 221, and performing some simple computations, we have

$$\begin{vmatrix} \bar{E} & \bar{F} \\ \bar{F} & \bar{G} \end{vmatrix} = \begin{vmatrix} E & F \\ F & G \end{vmatrix} \begin{vmatrix} \frac{\partial x}{\partial \bar{x}} & \frac{\partial x}{\partial \bar{y}} \\ \frac{\partial y}{\partial \bar{x}} & \frac{\partial y}{\partial \bar{y}} \end{vmatrix}^2 \quad (5)$$

If we form the analogue of (4) in the new coordinates and make use of (5) and of the formula for a change of variables in a multiple integral, we obtain

$$\begin{aligned} \iint_{\Omega} \sqrt{\bar{E}\bar{G} - \bar{F}^2} d\bar{x} d\bar{y} &= \iint_{\Omega} \sqrt{EG - F^2} \begin{vmatrix} \frac{\partial x}{\partial \bar{x}} & \frac{\partial x}{\partial \bar{y}} \\ \frac{\partial y}{\partial \bar{x}} & \frac{\partial y}{\partial \bar{y}} \end{vmatrix} d\bar{x} d\bar{y} \\ &= \iint_{\Omega} \sqrt{EG - F^2} dx dy \end{aligned}$$

showing that integral (4) remains invariant under a change of coordinates.

Now we shall prove that the region function $f(\Omega)$ defined by (4) satisfies conditions (a) and (b).

Let Ω and Ω' be two congruent regions. We have to show that $f(\Omega) = f(\Omega')$. Because Ω and Ω' are congruent, there is a motion of the plane under which Ω is superposed on Ω' . Suppose that under this motion the coordinate axes Ox, Oy assume the positions $O'x', O'y'$. Apart from the old Beltrami coordinate system x, y , consider a new system x', y' with axes $O'x'$ and $O'y'$, and suppose that

$$ds^2 = E dx^2 + 2F dx dy + G dy^2$$

and

$$ds^2 = E' dx'^2 + 2F' dx' dy' + G' dy'^2$$

are two representations for the metric form for the Lobachevskian plane in the old and new system, respectively. Let M be any point of Ω and M' the point into which M is mapped when Ω is superposed on Ω' . It is easily seen that the old coordinates of M are the same as the new coordinates of M' and the values of E, F, G at M are respectively equal to the values of E', F', G' at M' . Therefore

$$\int_{\Omega} \int \sqrt{EG - F^2} dx dy = \int_{\Omega'} \int \sqrt{E'G' - F'^2} dx' dy'$$

But we have seen that the value of integral (4) considered over some region does not depend on the coordinate system chosen; thus

$$\int_{\Omega'} \int \sqrt{E'G' - F'^2} dx' dy' = \int_{\Omega} \int \sqrt{EG - F^2} dx dy$$

whence

$$\int_{\Omega'} \int \sqrt{EG - F^2} dx dy = \int_{\Omega} \int \sqrt{EG - F^2} dx dy$$

We have thus shown that the function

$$f(\Omega) = \int_{\Omega} \int \sqrt{EG - F^2} dx dy$$

satisfies condition (a). That this function also satisfies condition (b) is a straightforward corollary of the additivity property of the integral: if the region Ω is divided into two regions Ω_1 and Ω_2 , then

$$\int_{\Omega} \int \sqrt{EG - F^2} dx dy = \int_{\Omega_1} \int \sqrt{EG - F^2} dx dy + \int_{\Omega_2} \int \sqrt{EG - F^2} dx dy$$

or

$$f(\Omega_1 + \Omega_2) = f(\Omega_1) + f(\Omega_2)$$

We agreed above to call the area of Ω the value of the positive function $f(\Omega)$ satisfying conditions (a) and (b). In accordance with this definition and the above theorem, the area of a region may be expressed by the formula

$$f(\Omega) = k \int_{\Omega} \int \sqrt{EG - F^2} dx dy \quad (6)$$

where k is a constant which is fixed by the choice of the unit of measurement of area.

The unit of measurement of area will now be put in a definite connection with that of length.

In Euclidean geometry the connection between the unit of areas and that of lengths is established by choosing the unit for area as a square whose side is the linear unit. Something similar will be done in Lobachevskian geometry.

Consider once again the horosphere Σ which touches the plane at the origin of the chosen Beltrami coordinate system; to the Beltrami coordinates (x, y) on the plane there correspond the Cartesian coordinates (x', y') on Σ .

Let Q' be a square on Σ (square in the sense of Euclidean geometry of the horosphere Σ) having one vertex at the origin, one side on the positive semiaxis Ox' , and another side on the positive semiaxis Oy' . Let a be the length of the side of this square. To Q' there corresponds on the plane a certain quadrilateral Q (explicitly, Q is the preimage of Q' under a mapping of the plane onto Σ , which was defined in § 216).

Denote by $S(Q')$ the Euclidean area of Q' [$S(Q') = a^2$] and by $S(Q)$ the area of the quadrilateral Q for a certain area unit on the plane. The unit of areas chosen will be subject to the condition

$$\lim_{a \rightarrow 0} \frac{S(Q)}{S(Q')} = 1$$

With this condition we shall find k in formula (6). Note that the (closed) region Q' on Σ is determined in the Cartesian coordinates (x', y') by the inequalities $0 \leq x' \leq a$, $0 \leq y' \leq a$. Because to the point (x', y') on Σ there corresponds on the plane a point with Beltrami coordinates $x = x'/R$, $y = y'/R$, it follows that the (closed) region Q is determined in Beltrami coordinates by the inequalities

$$0 \leq x \leq \frac{a}{R}, \quad 0 \leq y \leq \frac{a}{R}$$

Thus we have

$$S(Q) = k \int_0^{\frac{a}{R}} \int_0^{\frac{a}{R}} \sqrt{EG - F^2} dx dy = k \int_0^{\frac{a}{R}} \int_0^{\frac{a}{R}} \frac{R^2 dx dy}{(1 - x^2 - y^2)^{3/2}}$$

which yields, after elementary calculations,

$$\lim_{a \rightarrow 0} \frac{S(Q)}{S(Q')} = \lim_{a \rightarrow 0} \frac{k}{a^2} \int_0^{\frac{a}{R}} \int_0^{\frac{a}{R}} \frac{R^2 dx dy}{(1 - x^2 - y^2)^{3/2}} = k$$

Hence

$$k = 1$$

Thus with the unit of area chosen the area of any region Ω is given by the formula

$$S(\Omega) = \int_{\Omega} \int \sqrt{EG - F^2} dx dy$$

Notice that the area of a triangle Δ is then given by

$$S(\Delta) = R^2 D(\Delta)$$

where $D(\Delta)$ is the defect (this formula may be compared with formula (I'), § 48).

§ 223. Formulas

$$ds^2 = E dx^2 + 2F dx dy + G dy^2 \quad (\text{I})$$

$$\cos \varphi = \frac{E dx \delta x + F(dx \delta y + dy \delta x) + G dy \delta y}{\sqrt{E dx^2 + 2F dx dy + G dy^2} \sqrt{E \delta x^2 + 2F \delta x \delta y + G \delta y^2}} \quad (\text{II})$$

$$S(\Omega) = \int_{\Omega} \int \sqrt{EG - F^2} dx dy \quad (\text{III})$$

the first of which in the expanded form is

$$ds^2 = R^2 \frac{(1 - y^2) dx^2 + 2xy dx dy + (1 - x^2) dy^2}{(1 - x^2 - y^2)^2} \quad (*)$$

give in Beltrami coordinates the lengths, angles and areas on the Lobachevskian plane.

In their structure, these formulas completely coincide with formulas (I*), (II*), (III*) of § 215 giving the values of geometrical quantities on the Euclidean plane. Of course, the values of the coefficients E, F, G in formulas (I*)-(III*) of § 215 are different from those of the coefficients E, F, G in formulas (I)-(III) of the present article.

Because the quantities E, F, G in formulas (II) and (III) are the coefficients of form (*), the form (*) is said to define the metric for the Lobachevskian plane.

§ 224. Hitherto we have exclusively used Beltrami coordinates. Now the class of admissible coordinate systems will be enlarged. Starting with a given Beltrami coordinate system (x, y) , we introduce new coordinates by means of two relations of the form

$$u = u(x, y), \quad v = v(x, y) \quad (*)$$

provided that the functions $u(x, y), v(x, y)$ are continuously differentiable and have a nonvanishing Jacobian for all values of x, y such that $x^2 + y^2 < 1$. The quantities (u, v) will be treated as the new coordinates of a point $M(x, y)$. The conditions that the functions in question are continuously differentiable and that their Jacobian does not vanish are imposed so that the definition of a smooth curve given in § 220 for Beltrami systems remains valid in the

new coordinate system. Moreover, under these conditions, the equations (*) are reversible and their inversion yields the functions

$$x = x(u, v), \quad y = y(u, v) \quad (**)$$

which are continuously differentiable and which have a nonvanishing Jacobian. The direction of a smooth curve $u = u(t)$, $v = v(t)$ is given in the coordinates (u, v) by the ratio dv/du ; indeed, relations (**) imply that

$$\frac{dy}{dx} = \frac{\frac{\partial y}{\partial u} + \frac{\partial y}{\partial v} \frac{\partial v}{\partial u}}{\frac{\partial x}{\partial u} + \frac{\partial x}{\partial v} \frac{\partial v}{\partial u}}$$

and, hence, the direction parameter $k = dy/dx$ is known if the ratio dv/du is.

Writing formulas (I), (II), (III), § 223, in the new variables (u, v) yields formulas of exactly the same structure (but with different E, F, G):

$$ds^2 = E du^2 + 2F du dv + G dv^2 \quad (I)$$

$$\cos \varphi = \frac{E du \delta u + F (du \delta v + dv \delta u) + G dv \delta v}{\sqrt{E du^2 + 2F du dv + G dv^2} \sqrt{E \delta u^2 + 2F \delta u \delta v + G \delta v^2}} \quad (II)$$

$$S(D) = \int_D \sqrt{EG - F^2} du dv \quad (III)$$

which express the differential of an arc, the angle between curves and the area of a region in the system (u, v) . In order to use these formulas it is enough to know only the coefficients of the quadratic form (I), that is, $E = E(u, v)$, $F = F(u, v)$, $G = G(u, v)$. On account of this, the quadratic form (I) is said to define the metric for the Lobachevskian plane in (u, v) coordinates.

Let us examine one important example of the transformation of coordinates:

$$x = \tanh \frac{\xi}{R}, \quad y = \frac{\tanh(\eta/R)}{\cosh(\xi/R)} \quad (1)$$

where (ξ, η) are the new coordinates. If x, y satisfy the inequality $x^2 + y^2 < 1$, then equations (1) are uniquely reversible; accordingly, they define a transformation of coordinates on the entire Lobachevskian plane. Replacing x, y in the metric form

$$ds^2 = R^2 \frac{(1-y^2) dx^2 + 2xy dx dy + (1-x^2) dy^2}{(1-x^2-y^2)^2}$$

by the right-hand sides of relations (1) and performing some simple computations, we obtain the metric form for the Lobachevskian plane in ξ, η coordinates:

$$ds^2 = \cosh^2 \frac{\eta}{R} \cdot d\xi^2 + d\eta^2 \quad (2)$$

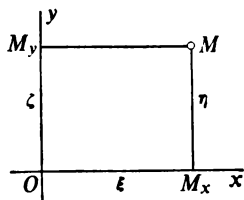


Figure 162

Therefore, formulas (II) and (III) yield

$$\cos \varphi = \frac{\cosh^2 (\eta/R) \cdot d\xi \delta\xi + d\eta \delta\eta}{\sqrt{\cosh^2 (\eta/R) \cdot d\xi^2 + d\eta^2} \sqrt{\cosh^2 (\eta/R) \cdot \delta\xi^2 + \delta\eta^2}}$$

$$S(D) = \int_D \cosh (\eta/R) \cdot d\xi d\eta$$

In metric form (2) the term with the product $d\xi d\eta$ does not appear. Notice that in the metric form

$$ds^2 = E du^2 + 2F du dv + G dv^2$$

in general (u, v) coordinates, F will vanish if and only if the family of coordinate curves

$$u = \text{const.}, \quad v = \text{const.}$$

is orthogonal. Indeed, the directions of coordinate curves are characterized by the differentials $dv, du = 0$ and $dv = 0, du$, with dv in the first case and du in the second being arbitrary variables. From this and (II) for the angle φ between the curves $u = \text{const.}, v = \text{const.}$, we have

$$\cos \varphi = \frac{F}{\sqrt{EG}}$$

Hence if $\varphi = \pi/2$, then $F = 0$, and conversely.

The absence of the term $d\xi d\eta$ in form (2) therefore implies the orthogonality of the family of coordinate curves $\xi = \text{const.}, \eta = \text{const.}$

Let us give a geometrical description for the coordinates ξ, η . Consider two mutually perpendicular axes Ox, Oy which are used to define the Beltrami coordinates x, y (Fig. 162). Taking an arbitrary point $M(x, y)$ in the plane, draw a perpendicular from M to Ox and denote the foot of this perpendicular by M_x . A comparison of the first of formulas (1), $x = \tanh (\xi/R)$, and the first of formulas (4) in § 218 reveals that they are identical. Accordingly, $\xi = OM_x$ and we can conclude that the equation $\xi = c$ (c is constant) represents a line perpendicular to the Ox -axis. For this curve formula (2) gives $ds^2 = d\eta^2$ or $ds = \pm d\eta$, the integration of which yields $M_x M = \pm \eta + a$ (a is constant). Setting $y = 0$ in the second of

the formulas (1), we find that $\eta = 0$. This implies that when M lies on the Ox -axis one must have $\eta = 0$. Accordingly, $a = 0$ and $M_x M = \pm \eta$. Notice finally that the second of formulas (1) shows that $\eta > 0$ if $y > 0$ and $\eta < 0$ if $y < 0$. Thus η expresses the signed segment $M_x M$ according to the usual convention. The quantities (ξ, η) are known as the *first coordinates* of M ; the quantities (ξ, ζ) , in terms of which the Beltrami coordinates (x, y) were expressed in § 218 (see also Fig. 162), the *second coordinates* of M . In Lobachevskian geometry always $\eta \neq \zeta$.

It is easy to understand that the coordinate curves $\xi = \text{const.}$ are lines perpendicular to the Ox -axis and the coordinate curves $\eta = \text{const.}$ are equidistants orthogonal to them.

8.4. Intrinsic geometry of a surface and the Beltrami problem

§ 225. By intrinsic geometry of a certain surface is meant the set of those properties of this surface which can be discovered by means of measurements that can be performed on the surface itself.

Euclidean planimetry is evidently a particular case of intrinsic geometry understood in the above sense.

The results obtained in the previous sections naturally lead to the question: Is it possible to treat Lobachevskian planimetry from the known point of view as an intrinsic geometry of a certain surface in the Euclidean space?

This question, raised in Beltrami's work "The Interpretation of Non-Euclidean Geometry" (1868), will be the subject of our discussion in the following subsections.

We begin with some simple facts of differential geometry. Although the majority (if not all) of them is well known, we act in this way in order to make our terminology clear and thus save the reader from possible misunderstanding.

First of all we shall make clear the precise meaning of the word "surface".

We shall confine ourselves to the most simple case of a surface without self-intersection which may be defined as a set of points of space (for the present the space is Euclidean).

Let a point set S be given in the Euclidean space. If M_0 is any point of S , then by a neighbourhood of M_0 in S we shall mean the subset $U(M_0)$ of this set which is the intersection of S with a neighbourhood of M_0 in the Euclidean space. The definition to follow requires the presence of neighbourhoods $U(M)$ of M having some definite properties.

In order to describe these properties, take an orthogonal Cartesian coordinate system with origin O and the axes Ox, Oy, Oz . Furthermore, imagine a plane with the two-dimensional Cartesian coordinate sys-

tem having axes u, v (henceforth this plane will be called the u, v -plane).

The set S will be called a surface if for any of its points M_0 there is a neighbourhood $U(M_0)$ such that its points have coordinates representable by the equations

$$\left. \begin{aligned} x &= x(u, v) \\ y &= y(u, v) \\ z &= z(u, v) \end{aligned} \right\} \quad (\alpha)$$

such that

(1) $x(u, v), y(u, v), z(u, v)$ are functions which are defined and single-valued in a region D of the u, v -plane;

(2) to every pair of numbers u, v belonging to D equations (α) assign a point with coordinates x, y, z belonging to the neighbourhood $U(M_0)$ (to distinct pairs of numbers u, v equations (α) assign distinct points, that is, equations (α) define a one-to-one correspondence between the points of D and the points of the neighbourhood $U(M_0)$);

(3) in D the functions $x(u, v), y(u, v), z(u, v)$ are continuous, have continuous first partial derivatives, and the rank of the matrix

$$\begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial y}{\partial u} & \frac{\partial z}{\partial u} \\ \frac{\partial x}{\partial v} & \frac{\partial y}{\partial v} & \frac{\partial z}{\partial v} \end{pmatrix} \quad (*)$$

is two.

The significance of the last condition will be explained later.

For the sake of visualization, it may be assumed without loss of generality that D is a simply connected region in the plane (u, v) . Then the corresponding neighbourhood $U(M_0)$ of any point M_0 will also be a simply connected region on the surface S .

Sometimes the neighbourhoods to which we have referred are called coordinate neighbourhoods. We shall not make our presentation more complicated by using this adjective, but by neighbourhoods of points of a surface below we shall mean neighbourhoods of this very form.

In some cases the entire surface is the neighbourhood of any of its points (for instance, a plane or a paraboloid). In the general case, a surface is the collection of a finite or infinite system of regions of the form mentioned above. Thus in defining a surface we allow the set of its points to have on the whole a complicated structure, but near every point its structure must be canonized in a definite respect.

In order to make the concept of a surface more convenient for application, it is useful to add to its definition one more connectedness condition. This condition may be stated, for example, in the following form.

Let U and V be neighbourhoods of two points of a surface. We shall say that these two neighbourhoods are connected by a chain of neighbourhoods if on the surface there are points and their neighbourhoods $U_1, U_2, \dots, U_n$ such that U_1 has a common part with U , U_n has a common part with V and U_k, U_{k+1} have a common part for any $k = 1, 2, \dots, n-1$. A surface will be called *connected* if any two neighbourhoods on this surface can be connected by a chain of neighbourhoods.

Consider a region U of the surface S that is represented by equations of form (α) . Every point M of U is determined from equations (α) by prescribing two numbers u and v . Therefore u, v will be called the coordinates of M on the surface and we shall use the notation $M(u, v)$, which is common in analytic geometry. These coordinates are often called *intrinsic*.

Starting with the u, v coordinates, infinitely many other intrinsic coordinate systems can be introduced in U . In order to do this, one need only form certain equations

$$\bar{u} = \bar{u}(u, v), \quad \bar{v} = \bar{v}(u, v)$$

which enable us to determine a new pair of numbers $\bar{u}, \bar{v}$ for every pair u, v ; the right sides of these equations must be subjected to exactly the same conditions stated in § 224 for equations $(*)$.

We defined a surface by means of three equations (α) . These equations can be replaced by one vector equation

$$\mathbf{r} = \mathbf{r}(u, v) \quad (\beta)$$

the left side of which contains the radius vector $\mathbf{r}$ of point M on the surface (that is, the vector $\overrightarrow{OM}$) and the right side contains the vector function with components $x(u, v), y(u, v), z(u, v)$.

If one uses equation (β) then the geometrical significance of condition (3) mentioned above in the definition of a surface can be easily explained. Namely, first of all the existence and continuity of the vectors

$$\mathbf{r}_u = \frac{\partial \mathbf{r}}{\partial u}, \quad \mathbf{r}_v = \frac{\partial \mathbf{r}}{\partial v}$$

are required and secondly that $\mathbf{r}_u \times \mathbf{r}_v \neq 0$, because the components of this vector product are the determinants of matrix $(*)$. The last result implies that the vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ are noncollinear; they then define a tangent plane to the surface.

Let us now consider equations of the form

$$u = u(t), \quad v = v(t)$$

These equations define a curve on the surface (the trajectory of the point $M(u, v)$ as t varies). The direction of this curve in the space is

given by the vector

$$\frac{d\mathbf{r}}{dt} = \mathbf{r}_u \frac{du}{dt} + \mathbf{r}_v \frac{dv}{dt}$$

The direction of the curve is evidently defined if the ratio of differentials, $du \div dv$, is given. Therefore $du \div dv$ will be referred to as the *direction parameter*.

Introducing the notation

$$\mathbf{r}_u^2 = E, \quad \mathbf{r}_u \mathbf{r}_v = F, \quad \mathbf{r}_v^2 = G$$

common in differential geometry we can find the square of the differential of an arc of the curve on the surface by setting

$$ds^2 = d\mathbf{r}^2 = (\mathbf{r}_u du + \mathbf{r}_v dv)^2 = E du^2 + 2F du dv + G dv^2$$

What is more, if $du \div dv$ and $\delta u \div \delta v$ are parameters of two directions corresponding to tangent vectors $d\mathbf{r}$ and $\delta\mathbf{r}$ to the surface, the angle φ between these directions is given by

$$\begin{aligned} \cos \varphi &= \frac{d\mathbf{r} \delta\mathbf{r}}{\sqrt{d\mathbf{r}^2} \sqrt{\delta\mathbf{r}^2}} = \frac{(\mathbf{r}_u du + \mathbf{r}_v dv)(\mathbf{r}_u \delta u + \mathbf{r}_v \delta v)}{ds \delta s} \\ &= \frac{E du \delta u + F (du \delta v + dv \delta u) + G dv \delta v}{\sqrt{E du^2 + 2F du dv + G dv^2} \sqrt{E \delta u^2 + 2F \delta u \delta v + G \delta v^2}} \end{aligned}$$

Finally, it is known from elementary analysis that if the region U on a surface corresponds to the region D on the plane (u, v) , then the area of U is given by the formula

$$\sigma = \int \int_{(D)} \sqrt{EG - F^2} du dv$$

Thus we have three basic relations:

$$ds^2 = E du^2 + 2F du dv + G dv^2 \quad (\text{I})$$

$$\cos \varphi = \frac{E du \delta u + F (du \delta v + dv \delta u) + G dv \delta v}{\sqrt{E du^2 + 2F du dv + G dv^2} \sqrt{E \delta u^2 + 2F \delta u \delta v + G \delta v^2}} \quad (\text{II})$$

$$\sigma = \int \int_{(D)} \sqrt{EG - F^2} du dv \quad (\text{III})$$

expressing the differential of an arc, the angle between two curves and the area of a region in the coordinates system u, v in terms of the functions $E(u, v)$, $F(u, v)$, $G(u, v)$. These formulas show that the lengths, angles and areas on a surface are completely determined by the coefficients of the quadratic form

$$ds^2 = E du^2 + 2F du dv + G dv^2 \quad (\text{I})$$

Therefore form (I) is said to define the metric of a surface and is known as *metric*.

It is clear that when the coordinate system is changed the coefficients of the metric form change accordingly as do the differentials of coordinates corresponding to the displacement of a point along the curve lying on the surface. If E, F, G are the coefficients of the metric form in one system of intrinsic coordinates and $\bar{E}, \bar{F}, \bar{G}$ are the coefficients in another system, and if du, dv and $\bar{d}u, \bar{d}v$ are the differentials of the old and new coordinates defined by one and the same element of the curve, then

$$E du^2 + 2F du dv + G dv^2 \equiv \bar{E} \bar{d}u^2 + 2\bar{F} \bar{d}u \bar{d}v + \bar{G} \bar{d}v^2$$

because the left and right sides represent the same quantity ds^2 .

If we know the formulas for transformation of coordinates and also E, F, G , then $\bar{E}, \bar{F}, \bar{G}$ can be easily found. Notice that the relation between E, F, G and $\bar{E}, \bar{F}, \bar{G}$ is obtained formally by algebraic computations. Therefore the computation of $\bar{E}, \bar{F}, \bar{G}$ when E, F, G are given can be performed by using formulas (5) of § 221, where the same question was settled algebraically.

If the coefficients of the forms are connected by relations (5), § 221, then we shall say that these two forms are mapped into each other by the transformation of coordinates. Such forms are called *equivalent*.

In light of this definition one can state that the metric of every surface in distinct intrinsic coordinates is defined by distinct metric forms, but all these forms are equivalent.

§ 226. Consider a region U on a surface S and a region U' on another surface S' . Assume that between the points of U and U' there is a one-to-one correspondence which is continuous in both directions. Then we shall also have a correspondence between the curves in U and the curves in U' ; namely, to every curve L in U there is assigned a curve L' in U' which is composed of points corresponding to the points of L . In exactly the same way, to every region V inside U there is assigned on U' a region V' composed of points corresponding to the points of V . A figure A' (for instance, a curve) in U' which corresponds to A in U will be called the *image* of A .

If every smooth arc l in U has for its image in U' a smooth arc l' having the same length as l , then the correspondence is called isometric or simply an isometry. The regions U and U' between which an isometric correspondence can be set up are called isometric to each other.

In order to obtain an analytic criterion for two regions to be isometric assume that intrinsic coordinates u, v have been introduced in U . In U' we shall introduce intrinsic coordinates in a special manner which will be related with the u, v system in U . Namely, to every point M' in U' we assign as their coordinates two numbers which are equal to coordinates (u, v) on U of that point M in U

which corresponds to M' . In other words, the coordinate system in U' is introduced in such a manner that the corresponding points of U and U' have numerically equal coordinates.

Let $E du^2 + 2F du dv + G dv^2$ and $E' du^2 + 2F' du dv + G' dv^2$ be the metric forms of U and U' in the given coordinates. Consider the respective elements of two curves in U and U' . They are characterized by the same differentials du and dv , and because of isometry we must have

$$E du^2 + 2F du dv + G dv^2 = E' du^2 + 2F' du dv + G' dv^2 \quad (*)$$

Since no conditions have been imposed on the pair of corresponding elements of two curves in U and U' , the quantities du and dv in $(*)$ are completely arbitrary. Accordingly, equation $(*)$ yields

$$E = E', \quad F = F', \quad G = G'$$

Therefore U and U' have the same metric forms in the given (u, v) -coordinates. The converse statement that if two surfaces have the same metric forms, then they are isometric is evident. (Notice that in arbitrary coordinates the metric forms of isometric surfaces may not coincide but they will be equivalent.)

From formulas (I)-(III) of § 225 it follows that in the case of isometry, besides the lengths of corresponding arcs being equal, the angles between corresponding directions and the areas of corresponding regions are also equal.

Therefore all those properties of a surface which may be obtained by performing measurements on this surface are identical for isometric surfaces. This enables us to say that isometric surfaces have a common intrinsic geometry. The intrinsic geometry which is common for a collection of isometric surfaces is defined by one metric form.

In order to see how to construct infinitely many distinct surfaces with a common intrinsic geometry, the reader should imagine that a surface is physically made of an elastic but inextensible material. Deform this surface so that there do not result any folds or cracks. Because the material is inextensible, the surfaces thus resulting are isometric among themselves and accordingly will have a common intrinsic geometry.

For example, by giving a cylindrical form to a sheet of paper, we can visually demonstrate that a part of the plane and a part of the cylinder have the same intrinsic geometry. However, if we try to put the sheet of the paper on a sphere or on a saddle (hyperbolic paraboloid), then in the first case folds appear while in the second case there appear cracks. This situation visually expresses the fact that the intrinsic geometry of every part of the sphere or the saddle is considerably different than geometry in any part of the plane.

The continuous deformation of a surface under which its intrinsic geometry is preserved is called *bending*.

Let us return to the subject-matter of § 224. There we defined the metric form of the Lobachevskian plane and obtained formulas (I)-(III) giving the lengths of curves, magnitudes of angles and areas of regions. The structure of these formulas is completely identical with that of formulas (I)-(III) of § 225. Therefore a natural question arises as to whether in the Euclidean space there is a surface whose metric form is equivalent to the metric form for the Lobachevskian plane. One may expect that the intrinsic geometry of such a surface coincides with Lobachevskian plane geometry, that is, it contains in the system of propositions all the axioms of Lobachevskian plane geometry.

If this question is formulated precisely in terms of isometry, then we can see that it leads to two different problems:

(1) Find a surface for each point of which there is a neighbourhood which is isometric to a certain region in the Lobachevskian plane.

One cannot yet assert about this surface that its geometry in the large is identical with the geometry of the Lobachevskian plane.

(For example, every point of a right circular cylinder has a neighbourhood which can be unfolded and superposed on a part of the Euclidean plane. But the geometry of a circular cylinder is in the large considerably different from the geometry of the plane.)

We shall say that on a surface which satisfies the conditions of the problem Lobachevskian geometry holds "in the small".

(2) Find a surface on which an isometry can be defined that maps this surface onto the entire Lobachevskian plane.

The intrinsic geometry of such a surface must be a realization of non-Euclidean plane geometry inside the Euclidean space. An affirmative answer to the second problem would have immediately implied the logical consistency of the two-dimensional non-Euclidean system. This was the very aim of Beltrami, to whom goes the credit, as mentioned above, for formulating these problems. But Beltrami solved only the first problem. As regards the second, it subsequently turned out that it has no solution. Namely, Hilbert showed that there is no surface in the Euclidean space with the desired property*.

The results of Beltrami, which are of geometric interest in their own right irrespective of the proof of the consistency of Lobachevskian geometry, will be presented in sufficient detail.

8.5. Geometry on a surface of constant curvature

§ 227. Our aim is to find, if possible, in the Euclidean space a surface for every point of which there exists a neighbourhood that is isometric to a region of the Lobachevskian plane.

* D. Hilbert: *Foundations of Geometry*, Appendix 5. It is worth pointing out that this theorem of Hilbert deals with surfaces whose current radius vectors are twice continuously differentiable with respect to the intrinsic coordinates (u, v) .

Assuming that such a surface exists, we denote it by S . We try to study the properties of S ; this will enable us to show the existence of such a surface along the way.

Let M_1, M_2 be any two points of S . According to the hypothesis, for each of these points there is a neighbourhood on S which is isometric to a part of the Lobachevskian plane. Denote the neighbourhoods of M_1, M_2 by U_1, U_2 , respectively. Assume that U_1 is isometrically mapped onto the region U'_1 in the Lobachevskian plane and M_1 is mapped into the point M'_1 inside U'_1 ; similarly, U'_2 denotes the region in the Lobachevskian plane obtained by an isometry of U_2 , and M'_2 in the point corresponding to M_2 under this isometry.

In the Lobachevskian plane, inside U'_1 there is another region $\overline{U}'_1$ which covers M'_1 and which is of such small dimensions that when this region is displaced congruently, under which M'_1 coincides with M'_2 , it covers a part of $\overline{U}'_2$ of the plane lying completely inside U'_2 . Furthermore, when $\overline{U}'_1$ is displaced to the new position $\overline{U}'_2$, any given direction at $\overline{M}'_1$ can be made to coincide with a given direction at M'_2 (this property of the set of motions was called in § 45 the transitivity of linear elements). Denote now by $\overline{U}_1$ and $\overline{U}_2$ the regions on S which correspond to $\overline{U}'_1$ and $\overline{U}'_2$ in the Lobachevskian plane under the isometry of U_1 and U_2 onto U'_1 and U'_2 .

Because the regions $\overline{U}'_1$ and $\overline{U}'_2$ are isometric, the regions $\overline{U}_1$ and $\overline{U}_2$ must also be isometric. In this way, given a point M_1 of S , there always exists a neighbourhood of this point which can be isometrically mapped onto a part of S such that M_1 is mapped into another preassigned point M_2 of the same surface. Moreover, the above discussions show that any given direction on S emanating from M_1 can be mapped into a given direction emanating from M_2 .

If the isometric regions of S are called congruent from the point of view of their intrinsic geometry and the terminology of § 45 is used, then the above result can be formulated in this form: the surface S admits of a set of motions which is transitive with respect to the linear elements.

Two features must be borne in mind:

(1) The isometric regions on S as images of the emanating Euclidean space are not congruent in general.

In the present case we are dealing with motions in the sense of the intrinsic geometry of the surface but by no means with motions in the sense of Euclidean geometry in space.

(2) The surface S in the large may not possess the property of moving about itself so freely as to make the set of motions transitive with respect to linear elements, even if they are considered from the point of view of intrinsic geometry.

In the present case we are dealing not with motions of the entire surface about itself but motions of sufficiently small pieces about the surface.

Even with these reservations one can discover a clear analogy between the surface S whose intrinsic geometry "in the small" is Lobachevskian and the surfaces on which elementary geometry is realized in the sense defined in § 45.

To visualize the motion in the sense of intrinsic geometry, imagine a piece of elastic but inextensible film placed tightly over the surface. The displacement of this piece along the surface represents a motion in the sense of intrinsic geometry if in every new position the piece being displaced is tightly affixed to the surface. The surface S in question must be bent in such a way that a piece of the inextensible elastic film affixed at any of its places could be freely displaced, without tearing off, along the surface and could rotate about any of its points; however, the length of the piece admitting of such a displacement may depend on the points between which the piece is displaced.

We confirm our discussion to the class of surfaces satisfying the additional condition of "third order smoothness". This means that the right sides of equations (α) of § 225 are assumed to be continuous triply differentiable functions. To such surfaces one may apply the classical theory of surfaces.

Taking into account Gauss' theorem on the invariance of the total curvature under isometries*, we can conclude on the basis of the above discussion that surface S has necessarily the same total curvature at all its points. Such a surface is known as a *surface of constant curvature*.

We shall now prove the following.

Theorem. *Every surface of constant curvature admits of a set of motions understood in the sense of intrinsic geometry which is transitive with respect to linear elements.*

We prepare the ground by performing some necessary computations. Let S be any surface of constant curvature. On this surface take a point M_0 and through this point draw a geodesic curve Γ . Starting at M_0 mark off an arc of length u on Γ , and through its final point draw a geodesic of length v perpendicular to Γ . Let the final point of the geodesic be denoted by M . In a neighbourhood $U(M_0)$ of M_0 the quantities u, v may be treated as the coordinates of M . Namely, u, v will be semigeodesic coordinates in the neighbourhood $U(M_0)$. In the u, v system the metric form is $ds^2 = E du^2 + dv^2$.

* The total curvature of the surface at a given point is the product of the principal curvatures at this point: $K = 1/R_1 R_2$. The proof of Gauss' theorem as well as other results of the theory of surfaces used in the present subsection can be found in A. V. Pogorelov: *Differential Geometry*, Noordhoff, Groningen, 1959.

We shall call Γ ($v = 0$) the basis geodesic of the coordinate system u, v and M_0 ($u = 0, v = 0$) the initial point or simply the origin.

Because the coordinate u is the length of an arc of the curve Γ , it follows that $ds^2 = du^2$ for $v = 0$. A comparison of this with $ds^2 = E du^2$, which results from the metric form when $v = 0$, yields

$$E(u, 0) \equiv 1$$

Notice further that since Γ is a geodesic, it follows that the geodesic curvature must vanish along Γ , that is, $1/\rho_g = 0$. Using the formula

$$\frac{1}{\rho_g} = \sqrt{EG - F^2} \left\{ \begin{matrix} 2 \\ 11 \end{matrix} \right\}$$

which is known from the theory of surfaces and which expresses the geodesic curvature of the coordinate curve $v = \text{const.}$, we have, noting that $1/\rho_g = 0$,

$$\left\{ \begin{matrix} 2 \\ 11 \end{matrix} \right\} = \frac{-FE_u + 2EF_u - EE_v}{2(EG - F^2)} = 0$$

for $v = 0$. But in the semigeodesic system $F(u, v) = 0$; thus the last relation yields

$$E_v(u, 0) = 0$$

We shall now determine the function $E(u, v)$ from the condition that the surface with the metric form

$$ds^2 = E du^2 + dv^2$$

has a constant total curvature.

It is known that the total curvature of a surface is given in the semigeodesic coordinates by the equation

$$K = -\frac{1}{\sqrt{E}} \frac{\partial^2 \sqrt{E}}{\partial v^2}$$

Thus we have to solve the differential equation

$$-\frac{\partial^2 \sqrt{E}}{\partial v^2} + K \sqrt{E} = 0 \quad (\alpha)$$

subject to the initial conditions

$$E(u, 0) = 1, \quad E_v(u, 0) = 0 \quad (\beta)$$

and K is constant.

We consider three cases.

(a) $K = 0$. The solution of (α) is

$$\sqrt{E} = \varphi(u)v + \psi(u)$$

Using the initial conditions (β) , we find $\psi(u) \equiv 1$ and $\varphi(u) \equiv 0$. Thus the metric form is of the form

$$ds^2 = du^2 + dv^2 \quad (\text{A})$$

(b) $K > 0$. Integrating equation (α) as a linear second-order equation, we have the general solution

$$\sqrt{E} = \varphi(u) \cos(\sqrt{K}v) + \psi(u) \sin(\sqrt{K}v)$$

In order to satisfy the initial conditions (β) the arbitrary functions of integration must be taken as $\varphi(u) \equiv 1$ and $\psi(u) \equiv 0$. Thus the metric form is

$$ds^2 = \cos^2(\sqrt{K}v) du^2 + dv^2 \quad (\text{B})$$

(c) $K < 0$. In this case the general solution of equation (α) is

$$\sqrt{E} = \varphi(u) e^{V^{-K}v} + \psi(u) e^{-V^{-K}v}$$

Using the initial conditions we have

$$\begin{aligned} \sqrt{E(u, 0)} &= \varphi(u) + \psi(u) \equiv 1 \\ (\sqrt{E(u, 0)})_v &= (\varphi(u) - \psi(u))\sqrt{-K} = 0 \end{aligned}$$

whence

$$\varphi(u) \equiv \psi(u) \equiv \frac{1}{2}$$

and

$$\sqrt{E} = \frac{e^{V^{-K}v} + e^{-V^{-K}v}}{2} = \cosh(V^{-K}v)$$

Thus the metric form is

$$ds^2 = \cosh^2(V^{-K}v) du^2 + dv^2 \quad (\text{C})$$

We thus find that in the semigeodesic coordinates with a basis geodesic the metric form of the surface of constant curvature K is defined uniquely only by the numerical value of K .

We now take any two points M_1 and M_2 of surface S and take each of them as the origin of a semigeodesic coordinate system. The directions of basis geodesics can be chosen in a totally arbitrary manner. Let U_1 be the region of existence of the semigeodesic system with M_1 as the origin and U_2 the region of existence of the semigeodesic system with M_2 as the origin.

If the positive number ε is sufficiently small, then for $-\varepsilon < u < +\varepsilon$, $-\varepsilon < v < +\varepsilon$ the point of the first system with coordinates (u, v) belongs to U_1 and the point of the second system with coordinates (u, v) belongs to U_2 .

Let Q_1 and Q_2 be the regions defined in the first and second coordinate systems, respectively, by the inequalities $-\varepsilon < u < +\varepsilon$,

$-\varepsilon < v < +\varepsilon$ (in their shape Q_1 and Q_2 are like squares). The computations just performed show that the metric form of Q_1 in the first coordinate system coincides with the metric form of Q_2 in the second. Therefore, if we set up a correspondence between the points of these two regions by equating the coordinates, then this correspondence will be an isometry. Therefore from the point of view of intrinsic geometry of the surface S the regions Q_1 and Q_2 are congruent. Since the choice of the basis geodesics in the coordinate systems used in the present argument is arbitrary, it follows that the set of congruent displacements on S is transitive with respect to the linear elements. ■

The investigation of the metric form of the surface of constant curvature enables us to state also the following theorem:

If two surfaces have the same constant curvature, then every sufficiently small part of one of them can be isometrically mapped onto a part of the other. Two surfaces having the same constant curvature have the same intrinsic geometry in small parts.

Note that two surfaces having distinct constant curvatures cannot be isometric. In fact, if these surfaces were to have the same metric forms in a certain coordinate system, then the computations of total curvatures of these surfaces would have resulted in the same constants.

§ 228. On the basis of the results so far presented, we may conclude that in the investigation of the intrinsic geometry in the small of a surface of a given constant curvature it is enough to study a certain representation of this class.

We consider three possible values for the total curvature $K = \text{const.}$: $K = 0$, $K > 0$ and $K < 0$.

(a) The most simple surface of constant zero curvature is the plane. The intrinsic geometry of the plane is Euclidean plane geometry. This geometry is defined by the metric form

$$ds^2 = du^2 + dv^2 \quad (*)$$

Since the metric form of any surface of zero curvature can be reduced to form (*), every sufficiently small part of such a surface can be isometrically mapped, or to use an alternative term, unfolded on the plane. On account of this the surfaces of zero curvature are called *developable*. At the same time, developable surfaces can be visualized as surfaces obtained in the process of bending the plane or a part of the plane or as surfaces formed by deforming parts of the plane.

For example, the parabolic cylinder is obtained by bending the entire plane. Its intrinsic geometry in the large is identical to Euclidean plane geometry.

The right circular cylinder is obtained by bending a plane strip. One must also join pairwise the points lying on the boundary of

this strip. In the small parts the circular cylinder has intrinsic Euclidean geometry, but its geometry in the large differs considerably from the geometry of the Euclidean plane.

The same thing can be said regarding a cone, which can be conveniently used to demonstrate motion in the sense of intrinsic geometry and to explain the significance of the conditions in theorems concerning this concept.

Let D be the part of a circular cone which is covered once by a plane disc with centre at M (the reader can imagine the cone in the form of a wooden model and the disc as made of paper). Every other part of the cone that can be covered by the same disc is isometric to D . Thus the motion of the disc about the cone is a motion in the sense of intrinsic geometry. That the motions in the sense of the intrinsic geometry of the cone are not identical to the motions in space is visually expressed by the deformation of the disc when it moves about the cone.

Displacing the disc, we can make its centre, originally at M , coincide with any point M' of the cone. However, if M' is assigned close to the vertex of the cone, then the size of the disc should be accordingly restricted. In any case, if the distance of M' from the vertex is less than the radius of the disc, then the disc does not find any place on the cone when the centre coincides with M' ; what is more, it should be noted that the part of the cone close to the vertex can be wrapped by the disc several times (therefore in theorems on motion on surfaces there figures the displacement of a sufficiently small part of it).

(b) The simplest surface having positive constant curvature $K > 0$ is the sphere with radius $R = 1/\sqrt{K}$.

Take the centre of the sphere at the origin of the orthogonal Cartesian coordinate system in space and introduce on the sphere intrinsic coordinates u, v by taking them equal to the geographical coordinates (that is, the longitude and latitude) multiplied by R . Every point on the sphere is determined by the equations

$$x = R \cos \frac{u}{R} \cos \frac{v}{R}$$

$$y = R \sin \frac{u}{R} \cos \frac{v}{R}$$

$$z = R \sin \frac{v}{R}$$

Then for any part of the sphere not containing the northern and southern poles for which $v = \pm 1/2 \pi R$, we have

$$ds^2 = dx^2 + dy^2 + dz^2 = \cos^2 \frac{v}{R} du^2 + dv^2$$

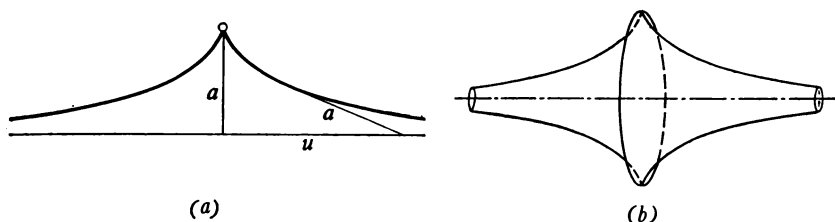


Figure 163

and setting $R = 1/\sqrt{K}$, we have

$$ds^2 = \cos^2 (\sqrt{K}v) du^2 + dv^2$$

which is exactly the same as expression (B) of the preceding subsection.

Thus the u, v system is a semigeodesic system for which the basis curve is the equator in the plane $z = 0$. By bending a part of the sphere we can obtain infinitely many other surfaces having positive constant curvature.

(c) One of the simplest surfaces with negative constant curvature $K < 0$ is the *pseudosphere*.

We shall now describe this surface. Consider a plane curve, called a *tractrix*, characterized by the following property: the length of the segment on the tangent to this curve from the point of contact to the point of intersection with a definite line is a constant.

In order not to spend time on long calculations, we ask the reader to find some of the properties of a tractrix without proof by using Fig. 163a, where this curve has been drawn.

In Fig. 163a the segment with constant length on the tangent line is denoted by a and the line along which the final point of the segment slides by u . It is clear that a tractrix has a point of return situated at a distance a from u ; this is the point of the tractrix farthest from u . From the point of return there issue two branches symmetrical to each other and each of which becomes closer to u without bound. This line is therefore an asymptote of the tractrix. It is also clear that at nonsingular points the tractrix has convexity towards the asymptote. A surface obtained by rotating the tractrix about the asymptote is called a *pseudosphere* (Fig. 163b).

A pseudosphere has two parts composed of regular points; each of these parts in stretching to infinity moves towards the axis of rotation. The two parts are joined to each other along the edge of return. In accordance with our definition of a surface (see § 225), we must assume the edge of return as not belonging to the surface. Henceforth, in speaking about a pseudosphere we shall mean one of its two regular parts. We will now show that at every point the pseudosphere

has a negative constant curvature. For this it is sufficient to prove that the curvature of a pseudosphere is a negative constant along a certain meridian of it.

Choose an orthogonal Cartesian system (x, y, z) such that the x -axis coincides with the axis of rotation of the pseudosphere and the plane $x = 0$ contains the edge of return. Consider the meridian of the pseudosphere located in the first quadrant of the xy -plane and assume that its equation is $y = f(x)$. For every $x > 0$ we have $a > y > 0$; what is more, since the point of the tractrix moves closer to the x -axis as x increases, it follows that $y' < 0$, and since the tractrix is convex towards the x -axis, we have $y'' > 0$.

Taking M to be any point of the meridian $y = f(x)$, we construct an outward normal to the pseudosphere at this point. Noting that the principal directions of the surface of rotation are the directions of its meridian and latitude, we shall compute the principal curvatures of the pseudosphere at M .

The normal to the pseudosphere is directed towards the concavity of the curve $y = f(x)$; therefore the principal curvature $1/R_1$ corresponding to the direction of the meridian is positive and is equal to the curvature of this meridian, that is,

$$\frac{1}{R_1} = \frac{y''}{(1 + y'^2)^{3/2}}$$

The curvature of the latitude is $1/y$; accordingly, the second principal curvature $1/R_2$ can be found by the formula

$$\frac{1}{R_2} = \frac{\cos \varphi}{y}$$

where φ is the angle between the normal and the segment y . This angle is evidently the angle of inclination of the tangent to the x -axis; accordingly, $\tan \varphi = y'$ and $\cos \varphi = -(1 + y'^2)^{-1/2}$. Hence

$$\frac{1}{R_2} = -\frac{1}{y \sqrt{1 + y'^2}}$$

The total curvature $K = 1/R_1 R_2$ at the points of the meridian $y = f(x)$ can now be expressed by the formula

$$K = -\frac{y''}{y(1 + y'^2)^2} \quad (1)$$

Draw a tangent to the curve $y = f(x)$ at the point $M(x, y)$ and denote by $(X, 0)$ the coordinates of the point of intersection of this tangent with the x -axis. Setting $Y = 0$ on the equation

$$Y - y = y'(X - x)$$

we find

$$X - y = -y/y'$$

By the definition of a tractrix,

$$X - x = -a \cos \varphi \quad (a = \text{const.})$$

Thus

$$y/y' = a \cos \varphi$$

Replacing $\cos \varphi$ by the expression $\cos \varphi = -(1 + y'^2)^{-1/2}$, we obtain the differential equation of the tractrix:

$$\frac{y \sqrt{1+y'^2}}{y'} = -a \quad (2)$$

whence

$$\begin{aligned} y'^2 (a^2 - y^2) &= y^2 \\ y'' (a^2 - y^2) &= y (1 + y'^2) \end{aligned}$$

The last two relations yield

$$y'' = \frac{y'^2 (1 + y'^2)}{y}$$

whence, noting (1), we get

$$K = -\frac{y'^2}{y^2 (1 + y'^2)}$$

Using equation (2), we finally have

$$K = -1/a^2$$

This shows that at all points the pseudosphere has the same negative curvature $-1/a^2$, where a is the parameter of the tractrix whose rotation generates the pseudosphere. Evidently, there exists a pseudosphere with some preassigned negative curvature. To obtain the meridian of the pseudosphere with a given curvature, one need only integrate equation (2) for the given value of the parameter a . In view of the previous result, we can assert that in the neighbourhood of any point of the pseudosphere the metric form in semigeodesic coordinates (with the basis geodesic) is of the form

$$ds^2 = \cosh^2 (\sqrt{-K}v) du^2 + dv^2$$

(formula (C), § 227).

By bending a piece of the pseudosphere, we can obtain infinitely many other surfaces with negative constant curvature.

Thus given any K ($-\infty < K < +\infty$), there exists a surface of constant curvature K in the Euclidean space.

As regards the solution to the Beltrami problem, we arrive at the following conclusion: if in the Euclidean space there exist surfaces on which Lobachevskian geometry in the small is realized, then one of these surfaces will be either a sphere or a plane or a pseudosphere.

Notice that semigeodesic coordinates are introduced on the surface in exactly the same way as the first coordinates in the Lobachevskian plane (see § 224). Therefore in semigeodesic coordinates the metric form of the surface with Lobachevskian intrinsic geometry must coincide with the metric form of the Lobachevskian plane in the first coordinates. At the end of § 224 the metric form of the Lobachevskian plane was expressed in terms of the first coordinates ξ, η :

$$ds^2 = \cosh^2 \frac{\eta}{R} \cdot d\xi^2 + d\eta^2 \quad (**)$$

It remains for us to compare this expression with the metric forms of a sphere, the plane and a pseudosphere, which, as is known, have in semigeodesic coordinates the forms

$$ds^2 = \cos^2(\sqrt{K}v) du^2 + dv^2$$

$$ds^2 = du^2 + dv^2$$

$$ds^2 = \cosh^2(\sqrt{-K}v) du^2 + dv^2$$

respectively. We see that (**) coincides with the last of these forms (at $1/R = \sqrt{-K}$).

Thus Beltrami's theorem follows:

In the neighbourhood of every point of a pseudosphere Lobachevskian geometry holds.

Let us cut the pseudosphere along a certain meridian; this will result in a simply connected region D bounded by the edge of return and the boundaries of the cut. Let D' be the region in the Lobachevskian plane that is isometric to D . We shall now give a description of D'^* in terms of Lobachevskian geometry. Since meridians are geodesics of the pseudosphere, it follows that under an isometry that maps D onto D' the meridians are mapped into a system of lines. This will be a system of lines parallel to each other in the sense of Lobachevsky. The last result follows from the asymptotic closeness of the meridians and hence of their images. The images of latitudes will evidently be orthogonal trajectories of the above system of parallel lines, that is, will be oricycles (precisely, arcs of oricycles).

Thus D' is bounded by two rays of the parabolic pencil (see § 39) and the arc of the oricycle orthogonal to this pencil. In Fig. 164, this region has been shaded.

Using an artificial device, one can realize a larger part of the Lobachevskian plane in the Euclidean space. To do this, imagine a countable set of equal and coincident pseudospheres. These pseudospheres are located in a definite order; namely, on the pseudosphere, denoted by A_0 there is superposed pseudosphere A_1 , on it pseudo-

* We should have shown the existence of D' , but this question is omitted from discussion.

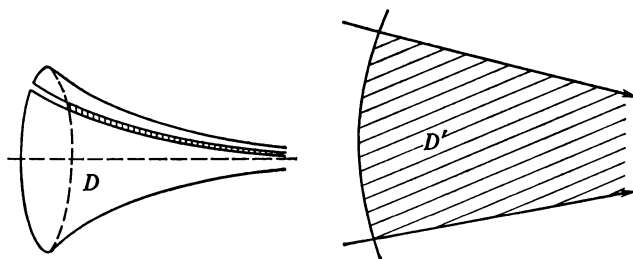


Figure 164

sphere A_2 , etc. Moreover, the pseudosphere A_0 itself is superposed on the pseudosphere A_{-1} , A_{-1} on A_{-2} , and so on. Now let us cut all the pseudospheres along a common meridian. For an observer who looks at the cut from the side of the axis one rim of the cut of every pseudosphere is the left side, while the other rim is the right side. Join the left rim of each A_n with the right rim of each A_{n+1} . Then a surface Σ is obtained that can be imagined as an infinite tape which is tightly wrapped around a pseudospherical mould. The surface Σ is evidently isometric to the part of the Lobachevskian plane which is situated on the side of concavity of a certain oricycle.

In other words, the part of the Lobachevskian plane that lies on the side of concavity of the oricycle can be realized in the Euclidean space in the form of an infinite wrapper (or, to use a common term, enveloping surface) of the pseudosphere.

As noted, Hilbert proved that in the Euclidean space there is no surface isometric to the entire Lobachevskian plane. Thus the attempt by Beltrami to realize non-Euclidean plane geometry in the form of intrinsic geometry of a certain surface could not but fail.

In spite of this fact, the researches of Beltrami are of a great principal value.

First of all, even a partial realization of non-Euclidean plane geometry in the Euclidean space changed the skeptical attitude of geometers towards the works of Lobachevsky. Therefore the discoveries of Beltrami played an important role in the general development of science.

Secondly, due to the works of Beltrami, Euclidean plane geometry, Lobachevskian plane geometry and the geometry on a sphere were united in a general differential-geometric setting. Namely, it was discovered that all these geometric systems are realized on surfaces of constant curvature K and correspond to the cases $K = 0$, $K < 0$ or $K > 0$.

What has been said makes clear the analytic source of the close relationship that exists between Lobachevskian geometry and spherical geometry.

Indeed, the metric form of the sphere

$$ds^2 = \cos^2 (\sqrt{\overline{K}v}) du^2 + dv^2 \quad (*)$$

and the metric form of the pseudosphere

$$ds^2 = \cosh^2 (\sqrt{-\overline{K}v}) du^2 + dv^2 \quad (**)$$

are distinct in the real region. But if $\sqrt{\overline{K}}$ or $\sqrt{-\overline{K}}$ are allowed to take imaginary values, then, as is known,

$$\cos (\sqrt{\overline{K}v}) = \cosh (\sqrt{-\overline{K}v})$$

Thus, if $\sqrt{\overline{K}}$ is replaced by $\sqrt{-\overline{K}}$, the forms (*) and (**) go over to each other.

8.6. Derivation of basic metric relations in Lobachevskian geometry

§ 229. In the present section we shall obtain a number of metric propositions for Lobachevskian geometry which were sidetracked from the main line of our presentation.

We shall not encounter any major difficulties. Having obtained the main metric formulas of Lobachevskian geometry (§§ 216-222), all other problems of a metric nature arising in this geometry can easily be solved by applying the formulas obtained.

As a basis of our computations, we shall take a system of Beltrami coordinates (x, y) . We know that (see § 216) Beltrami coordinates of any point in the Lobachevskian plane satisfy the inequality

$$x^2 + y^2 < 1 \quad (*)$$

Consider a Euclidean plane Σ with an orthogonal Cartesian coordinate system (x, y) . Inequality (*) defines on Σ the interior of a unit disc k_1 . To any point of the Lobachevskian plane having Beltrami coordinates (x, y) assign a point of Σ (lying in the interior of k_1) having for its Cartesian coordinates the same numbers x, y . In this manner we establish a special mapping of the entire Lobachevskian plane into the interior of k_1 ; under this mapping the images of Lobachevskian lines will be chords of k_1 (of course, with endpoints excluded).

Let us introduce an artificial metric in the interior of k_1 . Namely, for the distance between two interior points of k_1 take the number which is the distance between their preimages on the Lobachevskian plane, for the magnitude of the angle between two chords a and b take the number which is the magnitude of the angle between Lobachevskian lines that are preimages of the chords a and b ; the areas of regions are similarly defined. In practical terms this means that the computation of the main geometric quantities must be performed in Cartesian coordinates by using formulas of Lobachevskian geometry

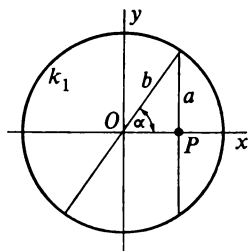


Figure 165

which express the respective formulas in terms of Beltrami coordinates.

In this way we obtain a realization of the Lobachevskian plane in the interior of the Euclidean disc k_1 . We shall be using this realization below.

It is worth noting that our conclusion will be of a general nature, that is, will not depend on the realization chosen. This is clear because relation (*) as well as the basic metric formulas of Lobachevskian geometry were derived by us from the axioms of Lobachevskian geometry irrespective of the objects on which these axioms are realized.*

§ 230. We shall now express the function $\Pi(l)$ in terms of elementary transcendental functions.

Let any line a and a point O at a distance $l > 0$ from this line be given in the Lobachevskian plane. From O draw the perpendicular OP to a and through O draw a line b parallel to a . The acute angle α between the lines b and OP is called the angle of parallelism for the segment $OP = l$ and is a function of l , that is, $\alpha = \Pi(l)$ (see § 33). We shall show that an extremely simple formula expresses $\Pi(l)$ in terms of elementary transcendental functions of l .

We take the origin of Beltrami coordinates at O and direct the Ox -axis along the segment OP ; point P will then have Beltrami coordinates $x = x_1$, $y = 0$. By the result of § 218, the line a is given by the equation $x = \tanh(l/R) = x_1 = \text{const.}$; accordingly, under the mapping of the objects of the Lobachevskian plane into the interior of k_1 , the line a is mapped into a chord perpendicular to the Ox -axis (Fig. 165), the line b is mapped into a chord which meets the chord a

* We cannot, however, claim to have proved that the system of axioms of two-dimensional Lobachevskian geometry is complete (for the concept of the completeness of a system of axioms see § 75). To do so, one should derive the basic metric formulas of Lobachevskian geometry without using axioms having a spatial character. Such a derivation was given by H. Liebmann: *Elementare Ableitung der nichteuklidischen Trigonometrie*, Berichte der Gesellschaft der Wiss. Leipzig, 1907. A simpler derivation was discovered by A. V. Pogorelov and is presented in his book *Osnovaniya Gheometrii* (Foundations of Geometry), Nauka, 1968 (in Russian).

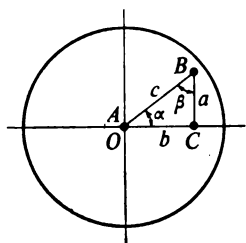


Figure 166

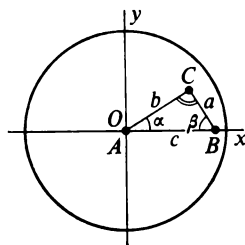


Figure 167

on the boundary of the disc k_1 (this follows from the definition of parallelism of lines in Lobachevskian geometry). It was shown in § 221 that the formula giving the angle between two directions at some point M in the Lobachevskian plane coincides with the Euclidean formula (II), § 215, provided M coincides with the origin. This means that the Euclidean angle between the chord b and the segment OP is α .

We have the Euclidean trigonometric relation $x_1 = \cos \alpha$. Besides this, we also have $x_1 = \tanh(l/R)$ (see the first of formulas (4) of § 218). The last two relations yield $\cos \alpha = \tanh(l/R)$, or after simple calculations, $\tan(\alpha/2) = e^{-l/R}$. Noting that $\alpha = \Pi(l)$, we obtain the formula of Lobachevsky:

$$\Pi(l) = 2 \arctan e^{-l/R}$$

§ 231. We shall now establish relations between sides and angles of a non-Euclidean triangle. First consider a right angled triangle ABC with legs $CB = a$, $CA = b$, hypotenuse $AB = c$ and with acute angles $CAB = \alpha$, $CBA = \beta$. Take the origin of Beltrami coordinates at A and the x -axis along the leg AC (Fig. 166). Let x_1, y_1 be the coordinates of C and x_2, y_2 of B . We have $x_1 = \tanh(b/R)$, $y_1 = 0$, $x_2 = x$, and we try to obtain y_2 from formula (3), § 217. Setting $\rho(B, C) = a$ in this formula, we have

$$a = \frac{R}{2} \ln \frac{1 - x_1^2 + y_2 \sqrt{1 - x_1^2}}{1 - x_1^2 - y_2 \sqrt{1 - x_1^2}} = \frac{R}{2} \ln \frac{1 + y_2 \cosh(b/R)}{1 - y_2 \cosh(b/R)}$$

Hence,

$$y_2 = \frac{\tanh(a/R)}{\cosh(b/R)}$$

Application of the Euclidean relation $y_2 = x_1 \tan \alpha$ to the triangle ABC , treating it as an object of Euclidean geometry, yields the following formula of Lobachevskian geometry:

$$\tanh \frac{a}{R} = \sinh \frac{b}{R} \tan \alpha \quad (1)$$

(a relationship between two legs and one acute angle).

Notice now that the Euclidean length c_e of the segment AB is expressed by the formula $c_e = \tanh(c/R)$ in terms of the length c of this segment in the sense of Lobachevskian geometry (for the proof of this, one has only to direct the x -axis from A along the segment AB and use the first of formulas (1), § 218). Using this fact, we see that the Euclidean formula $b_e = c_e \cos \alpha$ immediately yields a formula for Lobachevskian geometry:

$$\tanh \frac{b}{R} = \tanh \frac{c}{R} \cos \alpha \quad (2)$$

(a relationship between the hypotenuse, the leg and the adjacent acute angle).

Let the coordinate axes be now located with respect to the triangle ABC as in Fig. 167. We express the angle β by formula (3), § 221. First, for the coefficients E, F, G substitute the coordinates of B ; namely, $x = \tanh(c/R)$, $y = 0$ in formula (7), § 220. Then formula (3), § 221, yields

$$\cos \beta = \frac{dx \delta x + \left(1 - \tanh^2 \frac{c}{R}\right) dy \delta y}{\sqrt{dx^2 + \left(1 - \tanh^2 \frac{c}{R}\right) dy^2} \sqrt{\delta x^2 + \left(1 - \tanh^2 \frac{c}{R}\right) \delta y^2}}$$

Assuming that dx, dy correspond to the displacement along the Ox -axis ($dy = 0$) and $\delta x, \delta y$ to the displacement along the line BC ($\delta y/\delta x = -\tan \beta_e$, where β_e is the Euclidean magnitude of the angle ABC) and setting

$$1 - \tanh^2 \frac{c}{R} = \frac{1}{\cosh^2(c/R)}$$

we obtain from the last formula

$$\cos \beta = \frac{1}{\sqrt{1 + \frac{\tan^2 \beta_e}{\cosh^2(c/R)}}}$$

A comparison of this formula with the known formula $\cos \beta = (1 + \tan^2 \beta)^{-1/2}$ yields

$$\cosh \frac{c}{R} \tan \beta = \tan \beta_e \quad (*)$$

Notice that at the centre of the disc k_1 the Euclidean angles coincide with the Lobachevskian angles; therefore, $\alpha_e = \alpha$ and $\tan \beta_e = \cot \alpha_e = \cot \alpha$. This and formula $(*)$ yield a new relation for Lobachevskian trigonometry:

$$\cosh \frac{c}{R} \tan \beta = \cot \alpha \quad (3)$$

(a relationship between the hypotenuse and two acute angles).

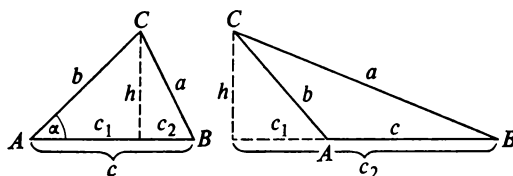


Figure 168

Formula (3) establishes a relationship between the linear magnitude and the angular magnitudes. There is no analogue of this formula in Euclidean geometry because there similarity of figures holds.

Let us go back to the triangle shown in Fig. 166. We have the Euclidean relation $c_e^2 = x_1^2 + y_2^2$, where $c_e = \tanh (c/R)$, $x_1 = \tanh (b/R)$ and $y_2 = \tanh (a/R)/\cosh (b/R)$. After some simple calculations this yields

$$\cosh \frac{c}{R} = \cosh \frac{a}{R} \cdot \cosh \frac{b}{R} \quad (4)$$

(a relationship between the hypotenuse and two legs).

We mention two more formulas which can easily be derived by the reader himself:

$$\sinh \frac{a}{R} = \sinh \frac{c}{R} \sin \alpha \quad (5)$$

(a relationship between the hypotenuse, a leg and the opposite acute angle) and

$$\cosh \frac{a}{R} \sin \beta = \cos \alpha \quad (6)$$

(a relationship between a leg and two acute angles).

§ 232. Let ABC now be any triangle in the Lobachevskian plane. If we draw the altitude h (as shown in Fig. 168) and use formula (4), we obtain

$$\cosh \frac{h}{R} = \frac{\cosh (b/R)}{\cosh (c_1/R)} = \frac{\cosh (a/R)}{\cosh (c_2/R)}$$

(the segments c_1 and c_2 have been shown in Fig. 168). This yields

$$\begin{aligned} \cosh \frac{a}{R} &= \frac{\cosh (b/R)}{\cosh (c_1/R)} \cosh \frac{c_2}{R} = \frac{\cosh (b/R)}{\cosh (c_1/R)} \cosh \left(\frac{c}{R} \mp \frac{c_1}{R} \right) \\ &= \cosh \frac{b}{R} \left(\cosh \frac{c}{R} \mp \sinh \frac{c}{R} \cdot \tanh \frac{c_1}{R} \right) \end{aligned}$$

But in view of (2) this gives

$$\tanh \frac{c_1}{R} = \pm \tanh \frac{b}{R} \cos \alpha$$

The last two results yield

$$\cosh \frac{a}{R} = \cosh \frac{b}{R} \cosh \frac{c}{R} - \sinh \frac{b}{R} \sinh \frac{c}{R} \cos \alpha \quad (\text{A})$$

§ 233. Formulas (1)-(6) as well as formula (A) were already established in § 61. However, these formulas were obtained only for a special model for Lobachevskian geometry. Here formulas (1)-(6) and (A) have been proved by using axioms of Lobachevskian geometry without making any assumptions regarding the nature of geometrical elements. Thus we have proved formulas (1)-(6) and (A) for any model for Lobachevskian geometry.

In § 62 we had compared formula (A) with the basic formula of spherical trigonometry

$$\cos \frac{a}{R} = \cos \frac{b}{R} \cos \frac{c}{R} + \sin \frac{b}{R} \sin \frac{c}{R} \cos \alpha \quad (\text{B})$$

Formula (B) becomes formula (A) if R is replaced by Ri ($i = \sqrt{-1}$). This implies that Lobachevskian trigonometry can be treated as trigonometry on a sphere having an imaginary radius. Such a relationship between Lobachevskian formulas and formulas of spherical trigonometry is explained comprehensively from the point of view of differential geometry. Lobachevskian geometry is a geometry of negative constant curvature $K = -1/R^2$, whereas geometry on a sphere is a geometry of positive constant curvature $K = 1/R^2$. When R is replaced by Ri , the metric form of a sphere becomes the metric form of the Lobachevskian plane. At the same time all the metric relationships of spherical geometry become the corresponding relationships of Lobachevskian geometry.

Chapter 9

Spatial Forms in Constant Curvature Geometry

9.1. Two-dimensional manifolds with differential-geometric metric

§ 234. We know that every surface in the Euclidean space has a completely defined intrinsic geometry. But, on its part, intrinsic geometry by no means defines the surface on which this geometry holds. Indeed, by means of bending one can obtain infinitely many surfaces that are distinct in their form but have a common intrinsic geometry.

Thus the structure of the spatial location of points of a surface is to a large extent coincidental for its intrinsic geometry. And in any case, if the metric form of a surface is known for a certain system of intrinsic coordinates, then all the results concerning the intrinsic geometry of this surface can be obtained without appealing to the embracing space. Therefore there arises the thought of generalizing the concept of intrinsic geometry so as to completely disengage it from the embracing space.

The benefit of such a generalization can be easily seen if one reviews the subject matter of the preceding section. We know that the metric of the Lobachevskian plane, just as the metric of every surface in the Euclidean space, is defined by a quadratic form. In an abstract form the Lobachevskian plane exists—it was proved at the end of Chapter 3. But it is not possible to superpose the entire Lobachevskian plane on a surface in the Euclidean space. For the given case, like many other geometric questions, the framework of the classical theory of surfaces turns out to be insufficient.

If we enlarge these frameworks, we arrive at such a wide concept of geometry that one scheme can include the most varied geometric systems, including the Lobachevskian system.

§ 235. Let a set R be given (the concrete nature of its elements is immaterial for us). The elements of this set will be called points and denoted by x, y, z , etc. Suppose that for every pair of points x, y there is defined their distance, which is a number $\rho(x, y)$. The set R together with the distance between its points will be called a *metric space* if the function $\rho(x, y)$ satisfies the following three conditions:

- (1) $\rho(x, x) = 0$
- (2) $\rho(x, y) = \rho(y, x) > 0$ for $x \neq y$
- (3) $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$

These conditions are known as the axioms of the metric space, and the last of these is called the triangle axiom.

Although these axioms demand very little of the function $\rho(x, y)$, they enable us to establish a number of important notions and theorems for any metric space. For example, one can define the concept of a *converging sequence of points* for any metric space: the sequence $x_1, x_2, \dots, x_n, \dots$ converges to a point a if $\rho(x_n, a) \rightarrow 0$. It can easily be shown that a sequence cannot converge to two distinct points. Indeed, suppose that $x_n \rightarrow a, x_n \rightarrow b$, where $a \neq b$. Since $\rho(a, b) > 0$, by the second axiom, it follows that for sufficiently large n , $\rho(x_n, a) < (1/2)\rho(a, b)$ and $\rho(x_n, b) < (1/2)\rho(a, b)$. Therefore $\rho(a, x_n) + \rho(x_n, b) < \rho(a, b)$, which contradicts the third axiom.

Next, one can define in a natural manner the concept of a continuous mapping of one metric space into another: the mapping $x' = f(x)$ of the space R into another space R' (that is, the correspondence of every point x of R with some point x' of R') is called *continuous* at a point a if every sequence $x_1, x_2, \dots, x_n, \dots$ converging to a is mapped into a sequence $f(x_1), f(x_2), \dots, f(x_n), \dots$ converging to the point $f(a)$: symbolically,

$$\lim f(x_n) = f(\lim x_n)$$

If the mapping $x' = f(x)$ is continuous at every point of R , then it is called continuous in the entire space, or simply continuous.

To cite examples of a metric space, we can mention in the first place the Euclidean plane (as well as the Euclidean space) and the Lobachevskian plane (as well as the Lobachevskian space). Next, consider a surface S in the Euclidean space (assuming it to be defined as in § 225). It can be shown that any two points x, y of S may be joined on this surface by a smooth or piecewise smooth arc; such an arc is necessarily rectifiable, that is, has a definite length. Now if for the intrinsic distance $\rho(x, y)$ between two points x, y of S we take the lower bound of the length of curves joining x, y on this surface, then $\rho(x, y)$ will satisfy Axioms 1-3 (the proof of this fact is omitted). Thus any surface with a definite intrinsic distance between its points is a metric space.

We find that the concept of a metric space embraces all the geometrical systems known to us. To underline how general this concept is we notice that on any set one can define a distance between pairs of points which satisfies Axioms 1-3.

Take any set M with elements $x, y, z, \dots$. Assume that $\rho(x, y) = 0$ if $x \equiv y$, $\rho(x, y) = 1$ if $x \not\equiv y$. Axioms 1-3 are then obviously satisfied, and therefore the set M with the distance defined is a metric space.

That the concept of a metric space is so general shows that for the development of a meaningful geometrical theory Axioms 1-3 are

inadequate. We shall now add a number of new, strong conditions to Axioms 1-3. This will result in a sufficiently concrete and at the same time very general class of metric spaces which will be called (two-dimensional) *Riemannian manifolds*, or *manifolds with a differential-geometric metric*.

These conditions are as follows.

(a) By an ε -neighbourhood ($\varepsilon > 0$) or simply a neighbourhood of a point a of a metric space we shall mean the set of all its points x for which $\rho(x, a) < \varepsilon$.

We require that for every point a of the space there exist a neighbourhood U that can be mapped into the Euclidean plane by means of a mapping which is one-to-one and continuous in both directions.

We introduce a coordinate system in U ; namely, by the coordinates u, v of a point x in U we mean the Cartesian coordinates of that point of the Euclidean plane which corresponds to x under the above mapping. The continuity condition imposed on this mapping means the following: if x_0 is a fixed point with coordinates u_0, v_0 and x a variable point with coordinates u, v , then $u \rightarrow u_0$ and $v \rightarrow v_0$ imply that $\rho(x, x_0) \rightarrow 0$, and, conversely $\rho(x, x_0) \rightarrow 0$ implies that $u \rightarrow u_0$ and $v \rightarrow v_0$.

(b) Let two neighbourhoods with coordinate systems defined in them have a common part. We require that in the common part of the two neighbourhoods coordinates of any point given in one neighbourhood can be uniquely expressed in terms of coordinates of this point considered as a point of the second neighbourhood by invertible equations the right sides of which have continuous partial derivatives and a nonvanishing Jacobian.

A metric space with these two conditions will be called a *two-dimensional smooth manifold*. For a smooth manifold the concepts of a smooth curve and direction can be defined.

By a *closed smooth arc* or briefly a *segment* in the neighbourhood U of a manifold S we shall mean the set of points of this neighbourhood the coordinates of which are given by the equations

$$u = u(t), \quad v = v(t)$$

where t lies in a closed interval $\alpha \leq t \leq \beta$ iff (i) for $\alpha \leq t \leq \beta$ the functions $u(t), v(t)$ are continuous and have continuous derivatives, (ii) for no t do the derivatives $u'(t), v'(t)$ vanish simultaneously, and (iii) for two different values of t the functions $u(t)$ and $v(t)$ do not both assume equal values simultaneously. The points of the segment corresponding to the values $t = \alpha$ and $t = \beta$ will be called its *endpoints*.

Obviously, the above properties of equations defining a segment are preserved when one changes to coordinates of any other neighbourhood containing this segment.

Thus because in the definition of a segment the choice of the neighbourhood covering this segment is immaterial, the concept of a segment is an invariant concept.

We shall say that at every point a smooth arc has a *direction* which is given by the ratio of differentials $du = u'(t) dt$, $dv = v'(t) dt$ (it is important here that $u'(t)$ and $v'(t)$ do not vanish simultaneously, because otherwise the ratio $du \div dv$ may become undefined). In a new coordinate system (u^*, v^*) the direction of the same curve is given by the ratio of the differentials

$$du^* = \frac{\partial u^*}{\partial u} du + \frac{\partial u^*}{\partial v} dv, \quad dv^* = \frac{\partial v^*}{\partial u} du + \frac{\partial v^*}{\partial v} dv$$

A finite system of smooth segments (lying, in general, in different neighbourhoods of the manifold) is called a *piecewise smooth arc* if in a suitable labelling of these segments the final point of the first coincides with the initial point of the second, the final point of the second coincides with the initial point of the third, etc. The initial point of the first segment and the final point of the last segment are called the endpoints of the piecewise smooth arc.

If at the common endpoints the adjacent segments have the same direction, then in this case the system of segments constitutes a *smooth arc*, which should be called *closed* because it has endpoints (which are the free endpoints of the first and last segments).

Similarly, one can define the concept of an open smooth arc and a piecewise smooth arc consisting of a countable number of segments i_n ($n = 0, \pm 1, \pm 2, \dots$) joined in such a way that the final point of the segment i_n coincides with the initial point of the segment i_{n+1} . An open smooth arc will simply be called a *curve*.

(c) One can define the concept of a rectifiable arc and its length in a metric space in the same way as is done in the Euclidean space.* We require that every fixed closed arc of any smooth curve $u = u(t)$, $v = v(t)$ be rectifiable and that the length of the arc on this smooth curve with one endpoint fixed and the other variable the function $(u(t), v(t))$ be a differentiable function of the parameter t .

(d) We finally require that in every neighbourhood with a given coordinate system (u, v) there exist three continuous functions $E =$

* Let a continuous arc L be given in space R , that is, let there be given a continuous image of the segment $\alpha \leq t \leq \beta$ the points of which (the image) are labelled by a suitable notation of their preimages from the segment $\alpha \leq t \leq \beta$ and are assumed to be ordered in increasing labels. Consider any system of points $\alpha = t_0 < t_1 < t_2 < \dots < t_{n-1} < t_n = \beta$ of the segment $\alpha \leq t \leq \beta$; to this system there corresponds a system of points $x_0, x_1, \dots, x_{n-1}, x_n$ on L . Form the sum $\sum_{i=0}^{n-1} \rho(x_i, x_{i+1})$. If the set of all sums (considered as a set of numbers) is bounded, then the arc L is said to be rectifiable; the upper bound of this set is the length of L .

$= E(u, v)$, $F = F(u, v)$, $G = G(u, v)$ by means of which the differential of the arc of the smooth curve $u = u(t)$, $v = v(t)$ is defined by the formula

$$ds^2 = E du^2 + 2F du dv + G dv^2, \quad \text{where } du = u'(t) dt, dv = v'(t) dt$$

A metric space satisfying all the above conditions is known as the *two-dimensional differential-geometric*, or *two-dimensional Riemannian, manifold*. By the way, to make this concept more convenient for application it is desirable to impose one more condition of connectedness besides those mentioned above; this can be stated in exactly the same form as the one for a surface (see § 225).

§ 236. The quantity φ defined by

$$\cos \varphi = \frac{E du \delta u + F (du \delta v + dv \delta u) + G dv \delta v}{\sqrt{E du^2 + 2F du dv + G dv^2} \sqrt{E \delta u^2 + 2F \delta u \delta v + G \delta v^2}}$$

will be called the angle between the directions $du \div dv$ and $\delta u \div \delta v$. The quantity on the right side is invariant under a change of coordinates (which can be proved as in § 224); accordingly, the coordinate system chosen in defining the angle is immaterial.

Finally, the area of a region D of a manifold S is defined as the value of the integral

$$\sigma = \iint_D \sqrt{EG - F^2} du dv$$

where the integral is taken over D . That the area is invariant was shown in § 222.

Thus lengths, angles and areas on any differential-geometric manifold are found by using exactly the same formulas as for a surface (see § 225, formula (I) (II), (III)).

Further, by the total curvature of a manifold at any point we can mean the quantity K which is found by the formula expressing the total curvature of a surface in terms of the coefficients of its metric form; the geodesic curvature of a curve can be similarly defined. However, in computing these quantities one has to differentiate twice the coefficients E, F, G of the metric form and the right sides of equations of the curve $u = u(t)$, $v = v(t)$. Therefore we must subject the concept of a smooth manifold and its metric to additional smoothness conditions; namely, we shall assume that the functions E, F, G are continuously differentiable up to the second order and the admissible transformations of coordinates up to the third order.

With these additional conditions a differential-geometric manifold is called *regular*.* Geodesics can be defined in the usual way on a regular manifold: either as extremals of the variational problem $\min \int ds$ or as curves with zero geodesic curvature. In order to find these geodesics we can use the known system of differential equations:

$$\frac{d^2u}{ds^2} + \left\{ \begin{matrix} 1 \\ 11 \end{matrix} \right\} \left(\frac{du}{ds} \right)^2 + 2 \left\{ \begin{matrix} 1 \\ 12 \end{matrix} \right\} \frac{du}{ds} \frac{dv}{ds} + \left\{ \begin{matrix} 1 \\ 22 \end{matrix} \right\} \left(\frac{dv}{ds} \right)^2 = 0$$

$$\frac{d^2v}{ds^2} + \left\{ \begin{matrix} 2 \\ 11 \end{matrix} \right\} \left(\frac{du}{ds} \right)^2 + 2 \left\{ \begin{matrix} 2 \\ 12 \end{matrix} \right\} \frac{du}{ds} \frac{dv}{ds} + \left\{ \begin{matrix} 2 \\ 22 \end{matrix} \right\} \left(\frac{dv}{ds} \right)^2 = 0$$

The family of geodesics coincides with the family of integral curves of this system. This implies that in a given direction only one geodesic passes through every point of a regular manifold.

For the geometry of manifolds the geodesics play the role of lines.

We have extended all the objects of intrinsic geometry of a surface to the case of a two-dimensional abstract differential-geometric manifold. Thus we have introduced a concept which is in any case wider than the concept of the intrinsic geometry of a surface in the Euclidean space because it embraces, for example, Lobachevskian plane geometry also.

We have defined a differential-geometric manifold by making use of the concept of a metric space with a number of additional conditions. These conditions were expressed in analytic terms; let us try to find their geometrical meaning.

Take any point $M_0(u_0, v_0)$ of the manifold with the metric form $E du^2 + 2F du dv + G dv^2$. Let E_0, F_0, G_0 be the values of E, F, G at M_0 . The following statements can be proved by performing computations, which will be omitted here.

Let $M(u, v)$ and $M'(u + \Delta u, v + \Delta v)$ be two points of the manifold and $\rho(M, M')$ their distance. If M and M' tend to M_0 , then

$$\lim \frac{\rho(M, M')}{\sqrt{E_0 \Delta u^2 + 2F_0 \Delta u \Delta v + G_0 \Delta v^2}} = 1 \quad (*)$$

In the Euclidean plane E we now take an oblique Cartesian coordinate system (u, v) whose scale vectors e_1, e_2 are chosen subject to the

* If we change to new coordinates $\bar{u} = \bar{u}(u, v)$, $\bar{v} = \bar{v}(u, v)$, the functions $\bar{E}, \bar{F}, \bar{G}$ in the new coordinates are expressed in terms of E, F, G and $\partial u / \partial \bar{u}$, $\partial u / \partial \bar{v}$, $\partial v / \partial \bar{u}$, $\partial v / \partial \bar{v}$ (see formulas (5), § 221); accordingly, the smoothness of functions $\bar{u}(u, v)$, $\bar{v}(u, v)$ must be higher by one than those of E, F, G .

conditions* that

$$|e_1| = \sqrt{E_0}, \quad |e_2| = \sqrt{G_0}, \quad \cos(e_1, e_2) = \frac{F_0}{\sqrt{E_0 G_0}}$$

If $M^*(u, v)$ and $M^*(u + \Delta u, v + \Delta v)$ are two points of the plane E , then their Euclidean distance $\rho_E(M^*, M^{*'})$ is given by the formula

$$\rho_E(M^*, M^{*'}) = \sqrt{E_0 \Delta u^2 + 2F_0 \Delta u \Delta v + G_0 \Delta v^2}$$

Using this result and (*) one can conclude the following.

For every point M_0 of the manifold there is a neighbourhood which can be mapped into the Euclidean plane E such that if M and M' are two points of the neighbourhood and $M^*, M^{*'} their images, then$

$$\rho(M, M') = \rho_E(M^*, M^{*'}) + \eta(M, M') \rho(M, M')$$

where $\eta(M, M')$ is an infinitesimal quantity together with $\rho(M_0, M)$ and $\rho(M_0, M')$.

In other words $\rho(M, M')$ differs from $\rho_E(M^*, M^{*'})$ by a quantity which is an infinitesimal of higher order in comparison to the dimensions of the neighbourhood.

This is the result which contains the main geometric meaning of conditions defining a Riemannian manifold. One can say that these conditions enable us to select a class of metric spaces which have Euclidean character in an infinitely small part.

The ideas presented here are well illustrated by the subject matter of §§ 220-222, where differential-geometric properties of the Lobachevskian plane were established by mapping it into the Euclidean plane in the form of a horosphere.

§ 237. Let a differential-geometric (Riemannian) manifold be given. Take any region H on it. It can be easily shown that H itself is a differential-geometric manifold. This manifold is a part of another wider manifold; if we were given only H , then we could "extend" it, that is, include it in another manifold, namely the original manifold.

In the investigation of intrinsic geometry of manifolds it is desirable to reject those which can be "extended", because otherwise such an investigation would get drowned in a mass of uninteresting relative frequencies.

Taking note of what has been said, we require the manifolds to be complete (which, however, is a far stronger condition than the non extensibility). The formulation of this condition requires the concept of a fundamental sequence; from the course of elementary analysis the reader is familiar with this concept for Euclidean geometry.

A sequence of points $x_1, x_2, \dots, x_n, \dots$ is called *fundamental* if $\rho(x_n, x_m)$ tends to zero as n and m increase without bound. In the

* These conditions are fulfilled because the positive-definiteness of the metric form implies that $E_0 > 0$, $G_0 > 0$, and $F_0^2 < E_0 G_0$.

Euclidean plane every fundamental sequence converges, that is, if $\lim_{\substack{n \rightarrow \infty \\ m \rightarrow \infty}} \rho(x_n, x_m) = 0$, then there exists a point x such that

$\lim_{n \rightarrow \infty} \rho(x_n, x) = 0$. In elementary analysis, this property of the

Euclidean plane is known as the general convergence principle.

For an arbitrary differential-geometric manifold the general convergence principle may not hold. To see this, it suffices to consider the following simple example.

In the Euclidean plane take a sequence of points $x_1, x_2, \dots, x_n, \dots$ which converges to the point x : $\lim_{n \rightarrow \infty} x_n = x$. This sequence is simultaneously a fundamental sequence. Let us delete the point x from the plane without changing the metric of the remaining part. In the manifold thus obtained the points $x_1, x_2, \dots, x_n, \dots$ still constitute a fundamental sequence, but now this sequence has no limit.

Differential-geometric manifolds in which there exist fundamental sequences without limits are called *incomplete*. A complete manifold cannot be extended.

We shall not examine incomplete manifolds below.

§ 238. By defining a two-dimensional differential-geometric manifold, we have obtained an important generalization of the concept of a surface and its intrinsic geometry. Let us explain its meaning and advantages.

In the classical surface theory, surfaces are regarded as objects of the Euclidean space. If a certain surface has the extrinsic equation

$$\mathbf{r} = \mathbf{r}(u, v)$$

then its intrinsic geometry is defined by the quadratic form $ds^2 = E du^2 + 2F du dv + G dv^2$ in which the coefficients

$$E = \mathbf{r}_u^2, \quad F = \mathbf{r}_u \cdot \mathbf{r}_v, \quad G = \mathbf{r}_v^2$$

are completely determined. Thus every surface has a definite metric or intrinsic geometry. In other words, the intrinsic geometry of a surface is defined by the peculiarities of location of its points in space. If using the theory of surfaces we wish to realize an intrinsic geometry with given properties, then we must look for a surface with the required intrinsic geometry, that is, a surface whose points are so located in space that the given metric just corresponds to this location. However, due to the nature of the problem of investigating the intrinsic geometry of a surface the spatial structure of a surface is not interesting.

When we examine the intrinsic geometry of an abstract manifold, which need not be thought of as being inside any space, we need not bother about the form of the surface as a spatial object.

We simplify our investigation by not having to be concerned with properties foreign to the figures under study.

On the other hand, the problem of finding a surface whose metric satisfies definite properties may not possess a solution, that is, there may not be a surface in the space with the given metric. But it may happen that this metric is realized on a certain abstract manifold, since the metric of an abstract manifold may be defined according to our convenience with a great degree of arbitrariness. The Beltrami problem can serve as a good example for this; many other examples will be given in the following pages.

The next section will be devoted to two-dimensional manifolds of constant curvature, that is, manifolds whose total curvature is the same at all points (the definition of total curvature of a manifold is given in § 236).

Because of their many properties, manifolds of constant curvature should be examined in the first place. It is enough to mention that only for these manifolds, among all the two-dimensional differential-geometric manifolds, sufficiently small pieces can be moved about. If two isometric regions of a manifold are treated as congruent, then the above statement can be formulated more precisely as follows.

Only manifolds of constant curvature admit of congruent displacements of their sufficiently small parts (that is, isometries of their parts into other parts) so that the collection of such displacements is transitive with respect to linear elements. To show this, the reader should go back to § 227, where exactly the same assertion was proved for a surface. Because the arguments of § 227 were presented exclusively for the concepts of the intrinsic geometry of a surface, they are directly applicable to metrized abstract manifolds.

What is more, on the basis of the results of § 227, we can assert that in the small parts of every manifold of constant curvature K Euclidean geometry is valid if $K = 0$, Lobachevskian geometry if $K < 0$ and spherical geometry if $K > 0$.

Thus abstract manifolds of constant curvature, just like surfaces, fall into three classes if considered in the small.

But if we examine these manifolds "in the large", we discover a wealth of properties, which would have been impossible to do within the framework of the elementary theory of surfaces.

9.2. Parabolic spatial forms

§ 239. A complete differential-geometric manifold of constant curvature is called the *spatial form of the geometry of a given constant curvature* (the completeness condition was stated in § 237). Two spatial forms of the geometry of a given constant curvature will be called equivalent if they are of the same topological type, that is,

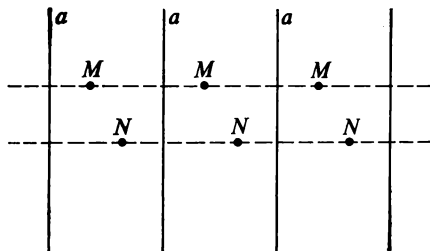


Figure 169

if they admit of a mapping of one into another which is one-to-one and continuous in both directions. Visually, this means that spatial forms can be deformed without “cracking” and “glueing”.

Such a condition results in a natural classification of all spatial forms in an easily visible set of classes of equivalent forms; every class is characterized by pointing out a certain representative.

Since we do not wish to give a complete topological classification of two-dimensional manifolds, we only mention that all two-dimensional manifolds can be classified as *open* or *closed*. A manifold is said to be closed if a convergent sequence can be chosen from any infinite set of its points (that is, if the Bolzano-Weierstrass principle holds for this manifold). A manifold which is not closed is said to be open.

Examples of closed manifolds are a sphere and a torus, while those of open manifolds are the plane, a region in the plane, or any region on a sphere which does not cover the entire sphere.

Henceforth, geometry of constant curvature K will be called *parabolic* if $K = 0$, *elliptic* if $K > 0$ and *hyperbolic* if $K < 0$.

§ 240. Among parabolic spatial forms the *Euclidean plane* must be mentioned in the first place (as examples of forms equivalent to it one may mention the parabolic cylinder, the connected part of the hyperbolic cylinder, a horosphere in the Lobachevskian space).

Apart from the Euclidean plane (and forms equivalent to it), there are four more spatial forms in parabolic geometry; their topological representatives are an ordinary cylinder, a one-sided cylinder, a torus and a one-sided torus. We shall examine them in this order.

We divide the plane into equal strips by a system of parallel lines $\{a\}$ (Fig. 169) and choose a direction in the plane; for example, the direction of the perpendicular to lines $\{a\}$ (the possibility of choosing direction along lines $\{a\}$ themselves is ruled out). In any strip take an arbitrary point M . Moving the strip selected in the given direction, it can be superposed on any other strip; the point M then assumes a new position which will again be denoted by M . The set of all points obtained in this manner from M will be denoted by $\{M\}$. Every set of points $\{M\}$ will be regarded as an element

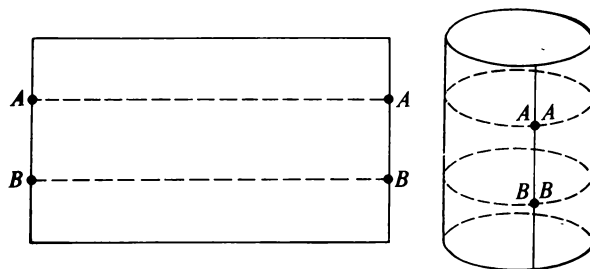


Figure 170

("point") of a new set R . We introduce a metric in R : if $x = \{M\}$ and $y = \{N\}$ be two points of R , then for their distance $\rho(x, y)$ we take the minimum of Euclidean distances between points of the set $\{M\}$ and the set $\{N\}$. By definition, $\rho(x, y) = \rho(y, x)$. Let us check if $\rho(x, y)$ satisfies all the axioms of a metric space (see § 235).

(1) If the sets $\{M\}$ and $\{N\}$ coincide, then the minimum Euclidean distance between points of $\{M\}$ and those of $\{N\}$ is zero; accordingly, $\rho(x, y) = 0$ for $x \equiv y$.

(2) If the sets $\{M\}$ and $\{N\}$ are distinct, then the minimum Euclidean distance between the points of $\{M\}$ and those of $\{N\}$ is greater than zero; this and the definition of $\rho(x, y)$ imply that $\rho(x, y) = \rho(y, x) > 0$ for $x \neq y$.

(3) Let $x = \{M\}$, $y = \{N\}$ and $z = \{P\}$ be any three points of R . Let N_1 be a point of $\{N\}$. There are points M_1 and P_1 in the sets $\{M\}$ and $\{P\}$, respectively, such that $\rho(x, y) = M_1N_1$ and $\rho(y, z) = N_1P_1$, where M_1N_1 and N_1P_1 are Euclidean distances. For Euclidean distances we have $M_1N_1 + N_1P_1 \geq M_1P_1$; but $M_1P_1 \geq \rho(x, z)$ and hence $\rho(x, y) + \rho(y, z) \geq \rho(x, z)$.

Thus R turns out to be a metric space.

The function $\rho(x, y)$ is defined in such a way that a sufficiently small ε -neighbourhood of any point $x = \{M\}$ of R is isometric to the ε -neighbourhood of the point M of the Euclidean plane; this means that R is a two-dimensional manifold with zero curvature.

Further, since the Euclidean plane is complete, R is also complete. Thus R has a certain parabolic spatial form.

In order to aid our discussion with visualization, imagine the plane as an infinite tape wrapped around a circular cylinder such that every strip wraps the cylinder just once. Then all the points of any set $\{M\}$ coincide with one point of the cylinder. Thus it is established visually that the circular cylinder is a spatial form of parabolic geometry equivalent to the manifold R examined before.

We shall arrive at the same conclusion if we simply identify all points of every set $\{M\}$ (in particular, points lying "opposite each

other" on various boundaries of a certain strip). It is clear that by joining pairwise the points lying on distinct boundaries of a strip (as shown in Fig. 170, where points joined have been denoted by the same letter) one obtains a "pipe".

Thus we have found a class of spatial forms of geometry of zero curvature the topological type of which is represented by a circular cylinder.

Let us go back to the plane divided by parallel lines $\{a\}$ into equal strips; take another line b perpendicular to lines $\{a\}$ (Fig. 171). Let M be any point in a certain strip. We move the chosen strip along b in a way such that it coincides with the adjacent strip; M will then have a new position. We reflect the point obtained in the line b and again denote the image by M . Continuation of the process gives an infinite set of points $\{M\}$.

Every set of points $\{M\}$ will be taken as an element of a metric space R : as in the preceding case, for the distance $\rho(x, y)$ between two points $x = \{M\}$ and $y = \{N\}$ of R take the minimum of Euclidean distances between the points of the set $\{M\}$ and those of $\{N\}$. It is easily seen that R has the spatial form of a geometry of zero curvature.

We shall obtain the same spatial form if by taking one strip we identify its boundary points lying in one set $\{M\}$ (the identification of points is shown in Fig. 172, where the points identified are denoted by the same letter). The manifold thus obtained is known as a *one-sided cylinder*.

Instead of an infinite strip of the plane, consider now the set of points lying in the interior of a rectangle $ABCD$ and the points that lie on its sides AB and CD excluding the points A, B, C, D themselves (thus the segments AD and BC are completely excluded). It can be easily proved that such a set is topologically equivalent to the infinite strip of the plane bounded by two parallel lines (that is, it can be mapped into this strip by a mapping which is one-to-one and continuous in both directions). Now identify pairwise the points of the segments AB and CD which are located symmetrically with respect to the centre of the rectangle (A is joined with C and B with D ;

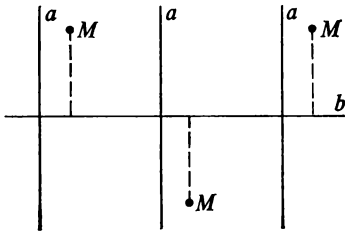


Figure 171

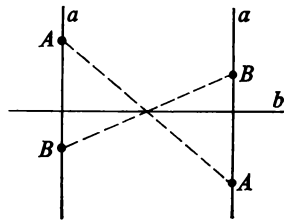


Figure 172

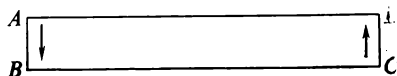


Figure 173

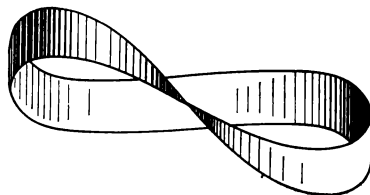


Figure 174

in Fig. 173 the arrows indicate the directions of AB and CD which must coincide after these segments have been joined). We thus obtain a manifold with Euclidean metric which is topologically equivalent to the one-sided cylinder (this, however, is not a spatial form because it is incomplete); it is called a *Möbius strip*.

In reality, we can join the opposite sides of the rectangle described above by taking a strip of paper and thus construct a model for the Möbius strip (Fig. 174). Using this model, one can easily see that the surface thus obtained is *one-sided*: it cannot be pointed in two colours so that the colours mix only at the boundary. The model for the Möbius strip to a certain extent makes clear our idea about a one-sided cylinder and also justifies this name.

Thus we have obtained a third class of parabolic spatial forms represented by a one-sided cylinder and topologically equivalent to the Möbius strip.

Consider now two systems of parallel lines, $\{a\}$ and $\{b\}$, in the plane which divide the plane into equal rectangles (Fig. 175). Take an arbitrary point M in one of them. By displacing the chosen rectangle in the directions of the lines $\{a\}$ and $\{b\}$, we can make it coincide with any other rectangle; the point M will then assume infinitely many new positions each of which will again be denoted by M . This results in an infinite set $\{M\}$; we shall treat it as an element of a new set R . In complete analogy with the preceding, we introduce a metric in R : if $x = \{M\}$ and $y = \{N\}$ are two points of R , then $\rho(x, y)$ is the minimum of the Euclidean distances between the points of $\{M\}$ and those of $\{N\}$. This metric space is a complete manifold with zero constant curvature, that is, a spatial form of parabolic geometry.

A visual representation of this form is given by the rectangle with the points of opposite sides being pairwise identified (Fig. 176, where arrows indicate directions of the sides which must coincide when these sides are identified; because all the vertices of the rectangle become one point, they have been denoted by the same letter). Notice now that when two opposite sides of the rectangle are identified, this results in a "pipe"; the identification of the other two sides yields a torus (Fig. 177).

Thus the topological type of the new class of spatial forms is represented by a torus (therefore they are called *ring-shaped*).

Let us go back to the plane divided into equal rectangles by lines $\{a\}$ and $\{b\}$, but now, in addition, draw a middle line between neighbouring lines $\{b\}$ in every strip (Fig. 178). Let M be any point in one of the rectangles. By moving the chosen rectangle along the strip between two lines $\{a\}$ and by superposing it successively on all other rectangles of this strip, we obtain infinitely many new positions for M ; all these points will be denoted by the same letter M . Now we move every rectangle of the strip in question along the lines $\{b\}$ into the adjacent strip and we reflect the point chosen in the midline of the rectangle (passing between the lines $\{b\}$); all the points thus obtained will again be denoted by M . This process will be continued indefinitely. The sets of points $\{M\}$ will be treated as elements of a new metric space R , where the metric is defined in exactly the same way as in all the preceding cases.

We thus obtain a parabolic spatial form represented by a rectangle in which the points of opposite sides are pairwise identified in the manner shown in Fig. 179. This manifold is known as a *one-sided torus*.

Let us try to make a model for the one-sided torus.

By joining two horizontal sides of the rectangle drawn in Fig. 179, we shall obtain a "pipe", but then in order to join the vertical sides of the rectangle as required by Fig. 179 we must pass one end of this pipe through the wall and join with the other end from within (Fig. 180). It is impossible to conceive a one-sided torus in the space in the form of a nonselfintersecting surface.

Parabolic spatial forms represented by a one-sided torus are called *one-sided ring-shaped forms*.

We have now established that there exist manifolds of zero curvature which are topologically equivalent to both an ordinary and a one-sided torus. This result was obtained only due to the fact that we introduced the concept of an abstract differential-geometric manifold. No regular surface in the Euclidean space having the topologi-

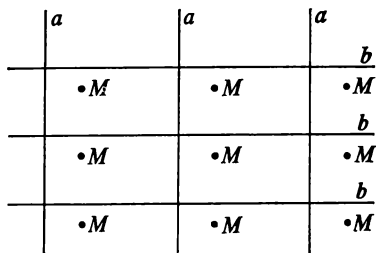


Figure 175

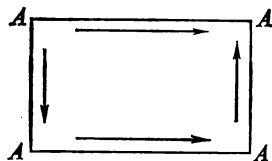


Figure 176

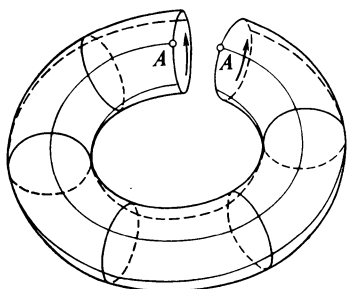


Figure 177

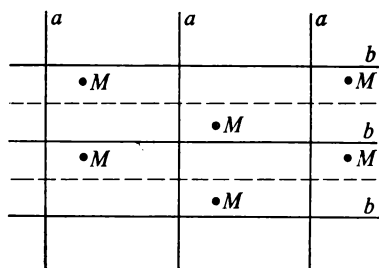


Figure 178

cal type of an ordinary or one-sided torus can have a natural metric of everywhere zero curvature. Thus we would not have been able to discover ring-shaped parabolic forms had we remained within the framework of the Euclidean surface theory.

§ 241. We have found five classes of spatial forms of parabolic geometry the topological representatives of which are

- (1) the plane,
- (2) the circular cylinder,
- (3) the one-sided cylinder,
- (4) the torus,
- (5) the one-sided torus.

Among these manifolds three are open (the first, second and third), while two are closed (the fourth and fifth); besides, there are three two-sided (the first, second and fourth) and two one-sided (the third and fifth). It is proved in topology that all these manifolds are topologically distinct.

Apart from the five manifolds mentioned, no other manifold can have parabolic geometry, that is, it cannot be metrized parabolically with the completeness condition being satisfied. The proof of this is sketched in Klein's *Vorlesungen über nicht-euklidische Geometrie*, Chapter 9. Of course, every region of the plane, cylinder, etc. is a metrized manifold with geometry of constant curvature, but the completeness condition is not satisfied in all these cases. By their definition, all parabolic forms have the same geometry in small parts as the Euclidean plane. But in the large every parabolic form has its own geometric system in which the role of points is played by elements of the manifold of the given form and that of lines by the geodesics of this manifold. The mutual relationships between points and lines are governed by all of Hilbert's axiom only for the geometric system of the first of the five parabolic forms. For the geometric systems of the remaining four forms completely different propositions turn out to be true, the majority of them being different from Euclidean propositions.

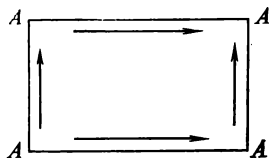


Figure 179

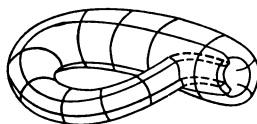


Figure 180

For example, in the geometry of a cylinder the geodesics of which are helices and circles orthogonal to the generators the statement that only one line passes through two points is false.

9.3. Elliptic spatial forms

§ 242. There are two classes of elliptic spatial forms; their representatives are (1) the sphere and (2) the elliptic plane.

The fact that the sphere is a spatial form of elliptic geometry can be immediately seen since the total curvature at all points of the sphere of radius r is $1/r^2$. Therefore the sphere, as a surface of the Euclidean space, has a metric of positive constant curvature.

We shall now show that there is a complete differential-geometric manifold of positive constant curvature, which is topologically not equivalent to the sphere.

Let a sphere S of radius r be given in the Euclidean space. Consider a set R the elements of which are pairs of antipodal points of S . We introduce a metric in R ; namely, if $x = \{M_1, M_2\}$ and $y = \{N_1, N_2\}$ are two elements of R (here M_1, M_2 and N_1, N_2 are two pairs of antipodal points of S), then for the distance $\rho(x, y)$ we take the minimum of the distances on S between the points of the pair $\{M_1, M_2\}$ and those of the pair $\{N_1, N_2\}$.

It can be easily shown (by using arguments analogous to the ones employed in the beginning of § 240) that $\rho(x, y)$ satisfies Axioms 1-3 of § 235, that is, R is a metric space.

Next, it is evident that for every point $x = \{M_1, M_2\}$ of R there exists an ε -neighbourhood which is isometric to the ε -neighbourhood of M_1 (or M_2) on S . Accordingly, R is a differential-geometric manifold of positive constant curvature. Finally, the completeness of S implies the completeness of R .

In this way, R is a spatial form of geometry of positive curvature. This form (as well as those equivalent to it) is called the *elliptic plane*.

The sphere and the elliptic plane are topologically not equivalent. To see this, notice that a sphere has the following property: every simple closed curve lying on the sphere divides the sphere into two

parts. This property must be preserved under a topological mapping of the sphere (that is, under a mapping of the sphere onto another manifold which is one-to-one and continuous in both directions). But the elliptic plane does not possess this property. Indeed, consider a set of pairs of antipodal points of a great circle of the sphere S ; call it L . The set L , being a subset of R , is topologically equivalent to the circle. Accordingly, L is a simple closed curve on the elliptic plane R . But L does not divide R into two parts because any two pairs of antipodal points of the sphere not belonging to L (that is, any two points of R not lying on L) can be mapped continuously into each other by bypassing L . This shows that the sphere and the elliptic plane are topologically distinct.

Thus we have found two spatial forms of geometry of positive curvature defining two distinct classes of forms; both these forms are distinct.

There are no other elliptic forms; however, the proof of this fact is not elementary and will be omitted.

§ 243. Here we shall describe two new representations for the elliptic plane.

(1) Let T be the set whose elements are all the lines passing through the centre of a sphere S (that is, T denotes a bundle of lines which is concentric with S). We introduce a metric in T by setting $\rho(x, y) = \alpha r$, where α is the smallest angle between the lines x and y and r the radius of S .

If with every line of T we associate a pair of antipodal points of S , where this line meets the sphere, we obtain an isometric mapping of T into R . This means that the bundle T with the metric introduced is a new representation of the elliptic plane.

(2) Let us complete the Euclidean space by adding elements at infinity in the same way as is done in projective geometry (see § 80). In space take any plane α and treat it as the projective plane, that is, having regard to points at infinity. We introduce a metric in the projective plane α .

To this end, take a bundle of lines T whose centre does not lie in α . With every line u of T associate a point x of α lying on u . This association is one-to-one (it is important here that the plane has been completed by adding points at infinity; due to this, to lines of the bundle which are parallel to α there correspond points at infinity of these lines). For the distance $\rho(x, y)$ between the two points x, y of α take the distance between those elements of T which correspond to the points x, y (the metric of T has been defined above). It is clear that with such a definition of distances on the plane α the correspondence between the points of α and the lines of T is isometric. Therefore the projective plane metrized in the above manner is a spatial form equivalent to T . We thus obtain a new representation for the elliptic plane in the form of the metrized projective plane.

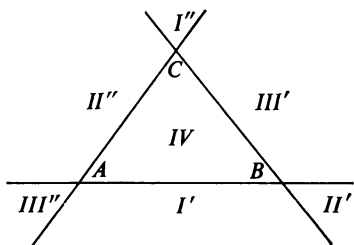


Figure 181

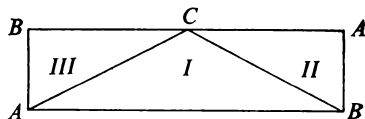


Figure 182

§ 244. Let us attempt to construct a topological model for the projective plane in the form of a surface in the Euclidean space topologically equivalent to it.

In the Euclidean plane α take three nonconcurrent lines. These lines divide α into seven regions, labelled in Fig. 181 by I' , I'' , II' , II'' , III' , III'' , IV . If the points at infinity are added to α , then the distinct regions I' and I'' are joined to give a connected region; it will be denoted by the roman numeral I and called a triangle because it is bounded by the segments of three lines. Similarly, II' and II'' are joined into one triangle II and III' and III'' into one triangle III . Thus the plane augmented by points at infinity is divided by the three lines into four triangles I , II , III , IV . Notice now that if triangle IV is deleted from the manifold, the remaining region will be topologically equivalent to the Möbius strip.

This will become visually clear if the triangles I , II , III are drawn as in Fig. 182. The reader can easily see that the scheme according to which the triangles I , II , III are joined in Fig. 182 is not different from the one used to join the triangles having the same numbers in Fig. 181. The vertices A , B of triangle II should be viewed as identified with the vertices denoted by the same letters for triangle III . It is evident that under such an identification triangles I , II , III constitute a Möbius strip the contour of which consists of straight links CA , AB , BC . The plane augmented by points at infinity is obtained by adjoining the contour of the Möbius strip to the contour of triangle IV .

It is known that a triangle is topologically equivalent to a part of the sphere that remains after a circular hole has been made in the sphere. Thus if the contour $CABC$ of the Möbius strip is jointed with the contour of triangle IV , this results in a surface which is topologically equivalent to the one that is obtained by gluing the Möbius strip to the sphere with one hole. This surface is one-sided. In the three-dimensional Euclidean space it is impossible to form such a surface without selfintersection. The projective plane viewed as a sphere with one hole which has been glued to the Möbius strip

enables us to visualize the peculiarities of the mutual location of the projective lines on the projective plane. This easily shows, for example, that a projective line does not divide the projective plane into two parts.

One can see this if a disc is cut out from the projective plane without touching the given line a ; the remaining part of the plane will be a Möbius strip containing the line a , which for visualization is assumed to coincide with the midline of this Möbius strip, but a closed cut of the Möbius strip along the midline does not divide it into two parts; this can be found using a simple model made of paper.

§ 245. To the two elliptic spatial forms there correspond two geometric systems: geometry on the sphere and geometry on the elliptic plane. The geometry on the elliptic plane is nothing but a two-dimensional Riemannian geometry (see §§ 63-67). On this account the elliptic plane is also referred to as the Riemannian plane.

9.4. Hyperbolic spatial forms

§ 246. In contrast to elliptic geometry, which has two classes of topologically equivalent spatial forms, and parabolic geometry, for which there are five classes, hyperbolic geometry can be realized, with the completeness condition being satisfied, on an infinite set of topologically distinct two-dimensional manifolds. Even among closed manifolds there are infinitely many such manifolds on which the metric of constant curvature $K < 0$ can be defined.

The results obtained in §§ 216-222 and §§ 227-228 enable us to assert that every plane in an axiomatically developed Lobachevskian space is a differential-geometric manifold of negative constant curvature; this manifold is complete, and the proof of this fact is exactly the same as that of the completeness of the Euclidean plane (one has to use the axioms of absolute geometry only). We can therefore assert that the Lobachevskian plane is one of the hyperbolic spatial forms. The class defined by it is characterized thus: its representatives are the complete differential-geometric manifolds of negative constant curvature which are topologically equivalent to the usual (Euclidean) plane. All these manifolds are known as *hyperbolic planes*.

§ 247. Assuming a certain hyperbolic plane given, we shall show how to construct infinitely many other hyperbolic forms. First we shall examine closed forms.

We shall employ a method for constructing two-dimensional closed manifolds known from elementary topology. For some time we shall discuss the topic from a purely topological point of view, that is, configurations will admit of any continuous deformations, although the metric properties might be violated in the process. Moreover, in

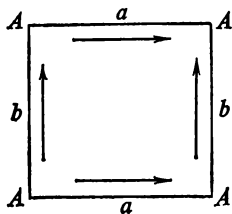


Figure 183

order to simplify the presentation we shall use visually descriptive devices.

Imagine a square made of a thin rubber band (Fig. 183). By joining its sides, denoted by a , so that the directions of these sides marked by arrows coincide, we transform this square into a pipe. Then by joining the boundaries of the holes of the pipe a torus is obtained (Fig. 184). Thus a torus can be regarded as a square in which the opposite sides are glued pairwise so that their directions, marked by arrows in Fig. 183, coincide and all the four vertices are joined to one point (in Fig. 183 the sides which are joined are denoted by one letter, and all four vertices have been denoted by one letter A ; in Fig. 184, where the torus has been shown, the notations have the same meaning as in Fig. 183).

Suppose that a hole is bored in the torus so that the torus becomes a "handle" (Fig. 185). Suppose that the boundary of this hole passes through A . Then on the original square the boundary of the hole is represented by a closed curve c that passes through A (Fig. 186). Tearing off the curve c at A and somewhat deforming the figure shown in Fig. 186, we can transform it into a pentagon shown in Fig. 187. Conversely, by gluing the sides, denoted by a , of this pentagon so that their directions marked by arrows coincide, and, similarly, by gluing the sides denoted by b and leaving the free side c as the boundary of the figure, we again obtain a handle.

If we join the boundaries of two handles, we get a "pretzel" (Fig. 188). At the same time, it may be evidently regarded as an octagon whose sides have been glued pairwise in the manner shown in Fig. 189, where the sides joined have been denoted by the same letters and their directions, which coincide, by arrows. Indeed, such an octagon results from two pentagons, representing handles, when they are joined along the free sides.

In the same way in which a torus, a surface of the first kind, is obtained by joining the sides of a square pairwise, and a "pretzel", a surface of the second kind, is obtained by joining pairwise the sides of an octagon, every two-sided closed surface of the p th kind may be obtained by joining pairwise the sides of a $4p$ -gon in a definite manner, which is shown in Figs. 190 and 191 for the particular case $p = 3$.

We shall now take up the construction of closed spatial forms of hyperbolic geometry. We shall show first of all that there exists a form which is topologically equivalent to a surface of kind $p = 2$ (that is, to a "pretzel").

In the Lobachevskian plane take a point O and draw through it four lines so that they form a proper star. Marking off congruent segments of length r on both sides from O on these lines and joining their ends by rectilinear segments, we obtain a proper octagon P_8 . We exclude from consideration those points of the Lobachevskian plane which are exterior to this octagon, and join its sides pairwise according to the scheme of Fig. 189. It is assumed that when the sides are joined, those points which divide these sides into the same ratio are also joined pairwise. Let R denote the set whose elements are (1) the interior points of the octagon P_8 , (2) the pairs of joined points of the sides and (3) the eight vertices (joined) of the octagon. Assuming that all points of the Lobachevskian plane representing some element x of R are denoted by one letter M , we shall express this element symbolically by $x = \{M\}$. The distance between two points P and Q of the Lobachevskian plane will be denoted by $d(P, Q)$.

We introduce a metric in R . Namely, if $x = \{M\}$, $y = \{N\}$ are two elements of R , then for the distance $\rho(x, y)$ take the smallest of the following two numbers:

$$d_1 = d(M, N), \quad d_2 = \min [d(M, T_1) + d(N, T_2)]$$

where T_1 and T_2 are any two identified points. Thus R is turned into a metric space which is topologically equivalent to the "pretzel". We can think of R as a "pretzel" on which the artificial metric has been defined by covering the "pretzel" by the octagon P_8 (that is, by means of the mapping of P_8 into the "pretzel"). This metric will be the usual Lobachevskian metric in every sufficiently small neighbourhood of any point of the "pretzel" except possibly the neighbourhood of that point at which all the vertices of the octagon coincide.

This point will be a singular point of the metrized "pretzel" if the sum of the interior angles of the octagon differs from four right angles.

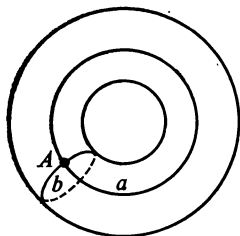


Figure 184

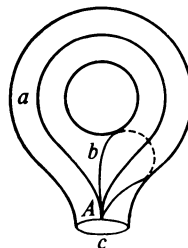


Figure 185

For the metric defined by us on the "pretzel" to be regular, we must make use of such an octagon the sum of the interior angles of which is equal to four right angles.

Thus the question of a regular hyperbolic metrization of the "pretzel" reduces to that of the existence in Lobachevskian geometry of a proper octagon with the desired angle-sum. This problem has a positive solution.

Indeed, let $S(r)$ denote the angle-sum of the octagon P_8 ; here r denotes the distance from the vertex of the octagon to its centre (this quantity figured in the construction of P_8). Notice that in the Lobachevskian plane the angle-sum of an infinitely decreasing triangle tends to π . Therefore $S(r) \rightarrow 6\pi$ as $r \rightarrow 0$ (that is, $S(r)$ tends to the angle-sum of the Euclidean octagon). This implies that $S(r_0) > 2\pi$ for a sufficiently small $r = r_0$. On the other hand, if $\beta(r)$ denotes the half of an angle of the octagon P_8 and $\delta = \delta(r)$ the half of the side of this octagon, then evidently $\beta(r) < \Pi(\delta(r))$, where Π is the Lobachevsky's function.

Since $\delta(r) \rightarrow \infty$ as $r \rightarrow \infty$ (see Lemma II, § 30) and $\Pi(\delta) \rightarrow 0$ as $\delta \rightarrow \infty$, the last inequality implies that $S(r) = 16\beta(r) \rightarrow 0$. Therefore $S(r) < 2\pi$ for sufficiently large r .

Since $S(r)$ is a continuous function, there is an $r = r_1$ such that $S(r_1) = 2\pi$, proving the desired result.

If one takes note of the fact that every quadrilateral in the Lobachevskian plane has an angle-sum less than 2π , it is easy to understand that it is impossible to construct by our method a hyperbolic metric which would be defined everywhere on a torus. It can be proved that the torus (a surface of kind $p = 1$) is not at all a topological representative of the spatial form of hyperbolic geometry.

In contrast, every two-sided closed surface of kind $p > 2$, like the case $p = 2$ just examined, can be hyperbolically metrized. This immediately follows from the fact that in the Lobachevskian plane there is a proper $4p$ -gon having the angle-sum 2π .

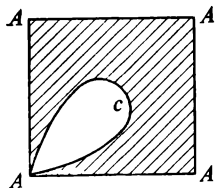


Figure 186

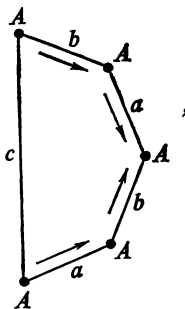


Figure 187

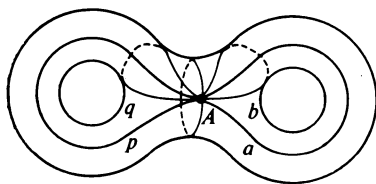


Figure 188

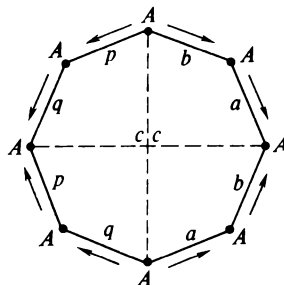


Figure 189

Analogous considerations enable us to establish that every one-sided closed surface, except the projective plane and the one-sided torus, also admits of hyperbolic metrization.* What has been said in the present subsection can be summed up in the following.

Theorem. *There are infinitely many distinct classes of closed hyperbolic spatial forms; their topological representatives are all closed surfaces except those which represent parabolic or elliptic spatial forms.*

§ 248. We shall say only a few words about open spatial forms of hyperbolic geometry. There are an infinite number of such forms. To make this clear, it is enough to mention some examples.

On the Lobachevskian plane, consider an infinite strip P_1 bounded by two parallel lines. Let P_2 be another exactly the same strip. If the lines bounding P_1 are identified with those bounding P_2 and the interior points of these strips are treated as distinct, then a manifold is obtained which is homeomorphic to the cylinder with a hyperbolic metric defined everywhere! on it; the completeness condition is evidently fulfilled. One may also act in a different manner: take two copies of the strip on the Lobachevskian plane bounded by two divergent lines, superpose one of them on the other and identify the coinciding boundaries. Then again a cylinder is obtained which can be metrized hyperbolically.

By the way, hyperbolically metrized manifolds obtained in two different ways have the same intrinsic geometry in the small, are of the same topological type, and are complete, but their metric properties in the large are essentially distinct (one of them is a pipe which expands indefinitely on one side and contracts indefinitely on the other side, while the second is a pipe that expands indefinitely

* One-sided closed surfaces are obtained from $2n$ -gons by identifying their sides pairwise by means of a special technique; for the particular case $n = 2$, this technique was demonstrated in § 240 in the construction of a one-sided torus. A brief discussion concerning the topology of closed surfaces sufficient for the present question, and also photographs of models of some one-sided closed surfaces in the Euclidean space (having selfintersection, of course) can be found in D. Hilbert and S. Cohn-Vossen: *Geometry and the Imagination*, Chelsea, New York, 1952.

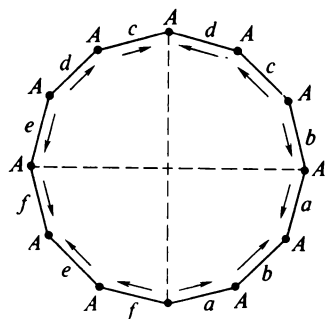


Figure 190

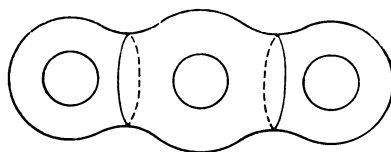


Figure 191

on both sides). If one considers two “triangles” with indefinitely extending sides which are superposed on each other (such figures obviously exist in the Lobachevskian plane) and identifies points on their boundaries, then a complete hyperbolically metrized open manifold of a new topological type is obtained.

This technique can be repeated without end. But in this way one cannot obtain, for example, a hyperbolic metric on the sphere; if we identify points on the boundary of two coincident discs in the Lobachevskian plane, we would obtain a sphere with hyperbolic metric but with a singular curve. In the present case our technique does not give a manifold with a metric defined everywhere, since the Lobachevskian plane (like the Euclidean plane) is not symmetrical with respect to the circle.

§ 249. Let us sum up the results of our investigation. We obtained an infinite number of different manifolds on which geometry of constant curvature is valid. All manifolds having a metric of the given constant curvature have a common geometry in the small. Each of them admits about them of congruent, in the sense of its geometry, displacements of their sufficiently small pieces, and the set of these displacements is transitive with respect to linear elements. But metrized manifolds of different topological types have different geometries in the large. To each of them there corresponds its own system of theorems expressing the properties of objects belonging to this manifold. The class of such geometries is a natural generalization of the geometries of Euclid, Lobachevsky and Riemann.

In the present section we studied exclusively two-dimensional geometries. A discussion of non-Euclidean geometries of dimension $n \geq 3$ can be found in Klein's *Vorlesungen über nicht-euklidische Geometrie*. See also P. K. Rashevsky: *Riemannian Geometry and Tensor Analysis*, Nauka, Moscow, 1967 (in Russian), and E. Cartan: *Leçons sur la géométrie des espaces de Riemann*, Gauthier-Villars, Paris, 1951.

Name Index

Aleksandrov, A. D. 9, 10
Archimedes, 15

Baldus, R. 139, 418
Beltrami, E. 40, 505, 511, 522
Bolyai, J. 35, 84

Cartan, E. 553
Cayley, A. 41
Chasles, M. 228
Chinn, W. G. 136
Cohn-Vossen, S. 551

Eddington, A. S. 471
Einstein, A. 457
Euclid 12, 33, 37, 82, 226

Gauss, C. F. 35, 40

Hilbert, D. 38, 42, 66, 206, 215, 216,
403, 511, 522, 551

Kagan, V. 33
Klein, F. 41, 187, 258, 346, 391, 395,
424, 543, 553
Kolmogorov, A. N. 212

Lambert, J. H. 21, 22, 23, 32, 33, 84

Legendre, A. M. 24, 32, 33, 84, 215.
Lie, M. S. 391
Liebmann, H. 524
Lobachevsky, N. I. 33, 34, 36, 37, 226.
Lorentz, H. A. 457

Minkowski, H. 457

Newton, I. 346
Nguyễn-Ca'hñ-Toan 10, 175

Pasch, M. 38
Pogorelov, A. V. 513, 524
Poincaré, J. H. 147, 149, 457
Poncelet, J. V. 226, 227, 228
Pontrjagin, L. S. 392

Rashevsky, P. K. 9, 10, 553
Riemann, G. F. B. 37, 40, 41. 226.

Saccheri, G. 21, 22, 23, 32, 33, 84
Staudt, K. G. C. von 231, 287
Steenrod, N. E. 136
Steiner, J. 228, 370

Vainshtein, I. A. 10

Zamorzaev, A. M. 10

Subject Index

- absolute conic
 - of Klein's group 422
 - of projective motion 411
- absolute geometry 84
- addition of elements 209
- adjacent angles 61
- affine geometry 401
 - n -dimensional 440
- affine group 400
 - basic invariants of 402
 - unimodular 404
- affine plane 400
- affine space, multidimensional 426
- affine transformation 400, 440
 - orthogonal 405
 - unimodular 403
- algebraic bundle 345
- algebraic curve 337
 - class of 344
 - order of 338
- algebraic pencil 337
 - class of 338
 - order of 344
- algebraic surface 345
- analytic representation for projectivity 309
- angle(s) 55
 - adjacent 61
 - congruence of 56
 - Euclid's definition of 13
 - exterior of 55
 - exterior point of 55
 - interior of 55
 - interior point of 55
 - magnitude of 75
 - non-Euclidean 156
 - of parallelism 101
 - sides of 55
 - vertex of 55
 - vertical 61
- angular unit 75
- annular surface 361
- antipodal points 178
- Archimedes' axiom 69
- Archimedes' postulates 15, 16
- area of triangle 136, 143, 405
- arithmetic space 427
- automorphism 400
 - Klein's 420
- axiomatic method in mathematics 223
 - and group theory 224
 - and topology 224
- axiomatic theory 188
- axiom of completeness 216
- axiom of parallelism 83, 84
- axioms of betweenness 45
 - projective 245
- axioms of congruence 53-5
- axioms of continuity 69, 70
 - projective 258
- axioms of elementary geometry, minimal system of 189, 205
 - completeness of 190, 216, 220
 - consistency of 189
- axioms of incidence 42, 43
 - projective 232
- axioms of projective geometry 232
- axis
 - of pencil 289
 - of perspective 234, 367
 - of rotation 68
- base of bundle 125
- basic invariant of affine group 402
- basis, orthonormalized 444
- basis of space 427
- basis points of pencil 386
- Beltrami coordinates 480
- Beltrami's theorem 521
- bending of surface 510
- betweenness, axioms of 45
 - projective axioms of 245
- Brianchon's theorem 377
- bundle 124, 298
 - algebraic 345
 - base of 125
 - centre of 124
 - class of 345
 - elliptic 124
 - parabolic 125

- canonical equations of a conic 353
- Cantor's axiom 70
- Cartesian realization of axioms 205
- central projection 226, 292
- centre of bundle 124, 298
- centre of perspective 234, 367
- class of algebraic curve 344
- class of algebraic pencil 338
- class of bundle 345
- coefficients of mapping 301
- complementary segments 247
- complete four-point 237
- completeness, axiom of 216
- complete system of axioms 222
- complex projective plane 343
- complex projective space 345
- cone, imaginary 437
 - interior of 437
 - null 360
 - ordinary 360
 - real nondegenerate quadric 437
- congruate triangles 58
- congruence, axioms of 53-5
- congruence of angles 56
- congruent images 130
- congruent segments 57
- congruent sets 66
- congruent systems of points 57
- conic 346
 - absolute 411
 - canonical equation of 353
- connected surface 507
- continuity, axiom of 69, 70, 258
 - of harmonic correspondence 252
- continuous mapping 530
- converging sequence of points 530
- coordinates 77-9, 427
 - intrinsic 507
- coordinate space 427
- coordinate systems 77-9
- correspondence of segments 66
- cross ratio of four elements 317, 318
- cross ratio of points 167
- curve(s), algebraic 337
 - degenerate 349
 - null 354
 - oval 359
- cyclic order of points 251
- cyclic points 422
- cylinder, one-sided 540

- d'Alembert's theorem 111
- Dedekind's axiom of continuity 259
- Dedekind's cut 80
- Dedekind's principle 79
 - in Euclidean geometry 150
- defect of a triangle 26, 136
- degenerate curve 349
- deleted projective line 259
 - linear order on 259
- Desargues' theorem 231, 241
- Desargues' first theorem 233
- Desargues' second theorem 234, 386, 387
- determinant of mapping 301
- developable surfaces 516
- diagonal points of four vertices 237
- differential-geometric manifolds (*see* Riemannian manifolds)
- differential geometry 40
- dihedral angle 104
 - edge of 104
 - faces of 104
 - linear angle of 104
- distance between events 463
- distance between points 407
- direction parameter 508
 - of line 493
- divergent lines 96, 105
- double points 382
- duality, principle of 325
- dual propositions 325

- elementary geometry 129
 - in Euclidean space 135
 - in Lobachevskian space 135
 - and projective geometry 230, 231
- elements
 - of the first and second kinds 298
 - harmonic sets of 231
- Elements* of Euclid 12ff
- elliptic bundle 124
- elliptic geometry 411
- elliptic involution 314
- elliptic metric 415
- elliptic plane 544
- elliptic projective motions 411
- elliptic spatial forms 544
- equation of a pencil 332fn
- equicomposed figures 136
- equidistant(s) 111
 - altitude of 112
 - base of 112
- equidistant surface(s) 121
 - altitude of 121
 - base of 121
- equivalent figures 394
- equivalent systems (*see* projectively equivalent systems)
- Erlangen Programm* 41, 395
- Euclidean geometry 39
 - and Dedekind's principle 150

- Euclidean geometry (*cont.*)
 - groups of objects in 38
 - and projective automorphisms 408
 - and Riemannian geometry 182
- Euclidean parallel postulate
 - and Lobachevskian geometry 189
- Euclidean plane 538
 - measurements on 474
 - metric of 477
- Euclidean space 39
 - metric form of 442
- Euclid's axioms 13
- Euclid's definitions 13
- Euclid's postulates 13
- Euler's formula 136
- exterior of angle 55
- fifth postulate 16-33
 - equivalent propositions 21
- figure(s) 394
 - equivalent 394
 - projection of 226
 - selfpolar 350
 - topologically equivalent 354
- first coordinates 505
- form(s), metric 508
 - valent 509
- four-point 237
 - complete 237
 - diagonal points of 237
 - opposite sides of 237
 - sides of 239
 - vertices of 237
- frame of reference 455
 - inertial 456
- fundamental sequence of points 535
- generalized numbers 209
- general Lorentz group 451
- general Lorentz transformation 451
 - geodesics 128
- geometrical elements 39, 42
- geometrical space 36
- geometrical properties of figures 395
- geometry, affine 401
 - axiomatic substantiation of 14
 - elliptic 411
 - and Euclid's Elements 12
 - and Greek scholars 12
 - of group 395
 - and group-theoretic principles 391
 - hyperbolic 411
 - parabolic 538
 - of projective group 395
- group(s) 391
 - affine 400
 - Klein's 420
 - n -member 397
 - orthogonal 406
 - subgroup of 392
 - of transformations 392
 - unimodular affine 404
- group properties 391
 - of projectivity 284
- group-theoretic principles and geometry 391
- half line 49
 - normalized parameters of 196
 - vertex of 49
- harmonically conjugate rays 240
- harmonic conjugate points 237
- harmonic correspondence, continuity of 252
- harmonic mapping 297
 - double points of 258
- harmonic pairs, division of 252
- harmonic sets of elements 231
- harmonic sets of points 240
- Hilbert's system of axioms, completeness of 221
- homogeneous coordinates 277, 280
- homogeneous coordinates, equation of a line in 279
- horosphere(s) 121
 - axis of 122
- hyperbolic geometry 411
- hyperbolic involution 314
- hyperbolic lines 415
- hyperbolic metric 415
- hyperbolic planes 547
- hyperbolic points 415
- hyperbolic projective motions 411
- hyperbolic spatial forms 547
- hyperplane in four-dimensional space 432
- identity transformation 393
- image of point 151
- images of second degree 346
- imaginary cone 437
- imaginary elements in geometry 342
- imaginary line 342
- imaginary point 342
- imaginary-unit vector 442
- incidence, axioms of 42, 43
- incomplete Riemannian manifolds 534
- index of Euclidean space 444
- integer points 261

- interior of angle 55
- interior points of angle 55
- interpretation of axioms (*see* realization of axioms)
- intrinsic coordinates 507
- invariant curves 111
- invariants of projection 227
- invariants of projective group 398
- inverse
 - of a circle 152
 - of a point 151
 - with respect to line 151
- inverse transformation 393
- inversion 151
- involution 300, 312
 - elliptic 314
 - hyperbolic 314
- isomorphic realization of axioms 221
- isotropic cone 443
- isotropic lines 443
- isotropic points 423
- isotropic vector 442

- Klein's automorphism 420
- Klein's group 420
 - absolute conic of 422

- Laguerre's formula 423
- length of segment 69, 70
- light cone 461
 - lower sheet of 461
 - upper sheet of 461
- limit circles (*see* oricycles)
- line(s) 42
 - in Cartesian realization 216
 - direction parameter of 493
 - Euclid's definition of 13
 - in four-dimensional space 431
 - hyperbolic 415
 - imaginary 342
 - at infinity 230
 - non-Euclidean 149
 - projective 230
- linear angle of a dihedral angle 104
- linear element 126
- linear-functional transformation 166
 - degenerate 167, 169
 - determinant of 166
 - nondegenerate 167, 169
- linearly dependent vectors 426
- linearly independent vectors 426
- linear mapping 301
 - coefficients of 301
 - determinant of 301
 - matrix of 301
- linear order 259
- linear space 426
- linear unit 70
- line conic(s) 346
 - theory of 356
- Lobachevskian geometry 33, 39
 - consistency of 189
 - and consistency of Euclidean geometry 148
 - definition of parallels in 96
 - logical consistency of 146
 - metric relations in 165, 523
 - and projective automorphisms 408
 - and Riemannian geometry 182
- Lobachevskian parallel axiom 84
- Lobachevskian plane, metric form for 489, 493
 - radius of curvature in 174
- Lobachevskian space 39
 - lines and planes in 104
- Lobachevskian trigonometry 175-8
- Lobachevsky's formula 174
- Lobachevsky's function $\Pi(x)$ 100-104
- Lorentz group 451
 - general 451
- Lorentz transformation 451
 - general 451

- magnitude of angle 75
- manifold(s), regular Riemannian 534
 - Riemannian 531
 - two-dimensional smooth 531
- mapping, axis of 297
 - centre of 297
 - coefficients of 301
 - continuous 530
 - determinant of 301
 - harmonic 297
 - linear 301
 - ordering 258
 - projective 380
- mathematics, axiomatic method in 223
- matrix of a linear mapping 301
- measuring 71
- metric of Euclidean plane 477
- metric form 508
 - of Euclidean space 442
- metric space 529
- midpoint of a segment 63
- Minkowski space 426, 445
- Möbius strip 354, 541
- model of a logical system 189
- motion(s) 66
 - product of 68
- multiplication of elements 209
- mutually congruent segments 57

- mutually congruent system of points 57
- mutually harmonic pairs of points 239
- mutually polar figures 350

- n*-dimensional Euclidean space 440
- n*-dimensional space 426
- n*-member group 397
- non-Archimedean system of numbers 209, 210
- nondegeneracy condition 442
- nondegenerate quadratic cone, real 437
- non-Euclidean angle 156
- non-Euclidean distance 171, 172
- non-Euclidean lines 149
- non-Euclidean points 149
- normalized parameters 196
- null cone 360
- null curve 354
- null surface 361

- one sided cylinder 540
- one-sided surface 541
- open spatial forms of hyperbolic geometry 551
- operation of projection 241
- operation of section 241
- opposite sides of four-point 237
- order of algebraic curve 338
- order of algebraic pencil 344
- ordered set 50
- ordering mapping 258
- order of surface 345
- ordinary cone 360
- oricycles 116
- orthogonal affine transformation 405
- orthogonal group 406
- orthogonal transformation 194
 - of first and second kinds 195
- orthonormalized basis 444
- oval curve 354
- oval surface 361

- pairs of planes, harmonically conjugate 244
- parabolic bundle 125
- parabolic geometry 538
- parabolic spatial forms 537
- parallel(s) 105
 - Lobachevsky's definition of 84
 - non-Euclidean theory of 84
 - transitive property of 95
- parallelism
 - in affine space 434
 - angle of 101
 - axiom of 82, 84
- parallel lines 18, 82, 90
- parallel planes 108
- parallel postulate (*see* fifth postulate)
- parallel translation 448
- parameters of projective group 397
- Pascal's theorem 377
- Pasch's axiom 45
 - in projective geometry 249
- pencil(s) 120, 228
 - algebraic 337
 - basis points of 386
 - with centre at point at infinity 228
 - equation of 332fn
 - of planes 289
- perspective, axis of 234, 367
 - centre of 234, 367
- perspective correspondence 367
- perspectivity 367
- $\Pi(x)$ 100-104
- plane(s) 42
 - affine 400
 - in four-dimensional space 431
 - hyperbolic 547
 - at infinity 230
 - parallel 108
 - projective 230
- plane geometry 126
- plane-quadratic 359
- Poincare's model 149, 165
- point(s) 42
 - in Cartesian realization 216
 - converging sequence of 530
 - cyclic 422
 - cyclic order of 251
 - Euclid's definition of 13
 - harmonic conjugate 237
 - harmonic set of 240
 - hyperbolic 415
 - imaginary 342
 - at infinity 114, 167, 184, 228
 - integer 266
 - isotropic 423
 - non-Euclidean 149
 - projection of 226
 - rational-binary 266
 - of space 394
- polar 347
- polar transform 350
- pole of line 349
- preimage of point 167
- "pretzel" (surface) 548, 549
- principle of duality 325
- product of elements 209

- product of motions 68
- product of transformations 393
- projection 241
 - central 226, 292
 - of figure 226
 - invariants of 227
 - of line on plane 101
 - of point 226
- projective axioms
 - of betweenness 245
 - of incidence 232
 - two-dimensional 326-8
- projective coordinate(s) 268
 - transformation formulae for 316
- projective correspondence 283, 292
- projective geometry 41, 226
 - axioms of 232
 - and elementary geometry 230, 231
- projective group 397
 - geometry of 395
 - invariants of 398
 - parameters of 397
- projective line 230
 - deleted 259
 - order of points on 244
 - topology on 251
- projectively equivalent bodies 300
- projectively equivalent figures 295
- projectively equivalent systems 280
- projective manifold, three-dimensional 298, 299
 - two-dimensional 298
- projective mapping 380
- projective metric 415
- projective motions 411
 - elliptic 411
 - hyperbolic 411
- projective plane 185, 230
 - complex 343
- projective properties 227
- projective scale, condensation of 263, 272, 273
 - integer points of 261
 - rational-binary points on 266
- projective space(s) 185, 230, 298
 - complex 345
- projective system of coordinates
 - on a line 258, 259
 - on a plane 270
 - in space 274
- projectivity 283, 289, 298, 390
 - analytic representation for 300, 309
 - of plane into plane 292
 - of space into space 299
- proper Euclidean space 445
- pseudo-Euclidean space 445
- pseudosphere 518
- Pythagorean theorem in Lobachevskian geometry 177
- quadratics 359
- quadric cone, real nondegenerate 437
- radius of curvature 174
 - in Riemannian geometry 183
- rational-binary points 266
- ray (*see* half line)
- rays, harmonically conjugate 240
- realization of axioms 188
 - Cartesian 205
 - isomorphic 221
- real nondegenerate quadric cone 437
- rectilinear generators of a surface 361
- regular Riemannian manifold 534
- Riemannian geometry 41, 178-187
 - and Euclidean geometry 182
 - and Lobachevskian geometry 182
 - and projective automorphisms 408
- Riemannian manifold(s) 531
 - incomplete 534
 - two-dimensional 533
- Riemannian plane 181
- rotation 68
 - axis of 68
- Saccheri's acute angle hypothesis 22
- Saccheri's obtuse angle hypothesis 22
- Saccheri's right angle hypothesis 22
- secant of equal inclination 93, 112
- second coordinates 505
- second-degree images 346
- section 241
- segment(s), complementary 247
 - length of 69, 70
 - midpoint of 63
 - mutually congruent 57
- selfpolar figure 350
- selfpolar three-point 350
- sequence of points, fundamental 535
- scalar product in affine space 440
- sides
 - of four-point 237
 - of three-point 233₁
 - of triangle 248
- signature of quadratic form 355
- smooth manifold, two-dimensional 531
- space 42, 394
 - arithmetic 427
 - of events 458
 - metric 529
 - Minkowski 426, 445

- n -dimensional 426
- projective 230
- proper Euclidean 445
- pseudo-Euclidean 445
- spatial forms, elliptic 544
 - hyperbolic 547
 - open 551
 - parabolic 537
- spherical geometry 127
- Staudt's theorem 287
- subgroup of group 392
- sum of elements 209
- superposition 66
- surface, algebraic 345
 - annular 361
 - bending of 510
 - connected 507
 - Euclid's definition of 13
 - null 361
 - one-sided 541
 - order of 345
 - oval 361
 - rectilinear generators of 361
- synchronization of clocks 456
- system of axioms, complete 222
 - projective 258, 259, 270
- system of points, mutually congruent 57
- tangent plane
 - to equidistant surface 124
 - to horosphere 124
 - to sphere 123
- theorem, Beltrami's 521
 - Brianchon's 377
 - Desargues' 231, 241
 - Desargues' first 233
 - Desargues' second 234, 386, 387
 - Pascal's 377
- theory of polars 346
 - mutuality principle in 348
- three-dimensional projective manifolds 298
- three-point 233
 - selfpolar 350
 - sides and vertices of 233
- topologically equivalent figures 354
- torus 542
- tractrix 518
- translation 68
- transformation(s), affine 400
 - automorphic 400
 - identity 393
 - orthogonal affine 405
 - unimodular affine 403
- transitive motions 126, 127
- triangle 248
 - angles of 136, 143
 - area of 405
 - defect of 26, 136
- triangulation 137
- trigonometry, Lobachevskian 175-8
- two-dimensional manifold 298
- two-dimensional projective axioms 326-8
- unimodular affine group 404
- unimodular affine transformation 403
- unit length 70
- unit vector 442
- valent form 509
- vector(s) 425
 - linearly dependent 426
 - linearly independent 426
- vertical angles 61
- vertices of four-point 237
- vertices of three-point 233

